Is Preference Alignment Always the Best Option to Enhance LLM-Based Translation? An Empirical Analysis

Hippolyte Gisserot-Boukhlef[1,4] **Ricardo Rei**[2] **Emmanuel Malherbe**[1]
Céline Hudelot[4] **Pierre Colombo**[3,4] **Nuno M. Guerreiro**[2,4,5,6]
[1]Artefact Research Center [2]Unbabel [3]Equall
[4]MICS, CentraleSupélec, Université Paris-Saclay [5]Instituto de Telecomunicações
[6]Instituto Superior Técnico & Universidade de Lisboa (Lisbon ELLIS Unit)
hippolyte.gisserot-boukhlef@centralesupelec.fr

Abstract

Neural metrics for machine translation (MT) evaluation have become increasingly prominent due to their superior correlation with human judgments compared to traditional lexical metrics. Researchers have therefore utilized neural metrics through quality-informed decoding strategies, achieving better results than likelihood-based methods. With the rise of Large Language Models (LLMs), preference-based alignment techniques have gained attention for their potential to enhance translation quality by optimizing model weights directly on preferences induced by quality estimators. This study focuses on Contrastive Preference Optimization (CPO) and conducts extensive experiments to evaluate the impact of preference-based alignment on translation quality. Our findings indicate that while CPO consistently outperforms Supervised Fine-Tuning (SFT) on high-quality data with regard to the alignment metric, it may lead to instability across downstream evaluation metrics, particularly between neural and lexical ones. Additionally, we demonstrate that relying solely on the base model for generating candidate translations achieves performance comparable to using multiple external systems, while ensuring better consistency across downstream metrics.[1]

1 Introduction

Neural metrics for machine translation evaluation that are trained to mimic human preferences, such as BLEURT (Sellam et al., 2020), COMET (Rei et al., 2020, 2022a), or Metric-X (Juraska et al., 2023), have become increasingly prevalent. These metrics offer greater accuracy and better reflect human judgments compared to traditional lexical metrics (Mathur et al., 2020; Kocmi et al., 2021; Freitag et al., 2022b; Kocmi et al., 2024) like BLEU (Papineni et al., 2002), METEOR (Banerjee and Lavie, 2005) or chrF (Popović, 2015), which mainly consider lexical overlap with a reference text. As such, researchers have attempted to leverage these improvements by integrating them directly into translation systems.

One appealing strategy to incorporate quality information to improve downstream translation performance involves using decoding strategies such as N-Best reranking and Minimum Bayes Risk (MBR) decoding (Kumar and Byrne, 2002, 2004; Eikema and Aziz, 2020; Fernandes et al., 2022; Freitag et al., 2022a). These techniques rely on generating multiple candidates to maximize a given quality metric at inference time, and research has shown that they consistently yield better results than likelihood-based decoding techniques (Eikema and Aziz, 2020; Koehn and Knowles, 2017; Ott et al., 2018).

With the rise of decoder-only LLMs in MT, quality-informed fine-tuning techniques have gained significant attention. Unlike decoding-based methods that inject quality information at inference time, fine-tuning modifies model weights using training sets induced with quality information. These approaches include filtering parallel training data based on a quality metric (Alves et al., 2024), distilling gains from more expensive quality-aware methods such as MBR (Finkelstein et al., 2024), or employing preference-based alignment techniques (Rafailov et al., 2024; Xu et al., 2024a), where the model learns preferences induced by quality metrics between candidate translations typically generated by multiple systems. In this work, we focus specifically on the latter.

Alignment techniques represent a paradigm shift from quality-aware inference time approaches, as they optimize the metric of interest *indirectly*. Understanding the impact of these approaches on translation quality is thus a relevant problem. While some studies have examined quality-

[1]All relevant preference datasets and aligned models, along with detailed evaluation metrics, are available at https://huggingface.co/collections/artefactory/translation-alignment-analysis.

Proceedings of the Ninth Conference on Machine Translation, pages 1373–1392
November 15-16, 2024

informed decoding techniques and their influence on translation output (Amrhein and Sennrich, 2022), there is still a gap in our understanding of how preference-based fine-tuning affects translation quality.

In this work, we aim to bridge this gap by examining the properties of preference-based alignment techniques, with a particular focus on Contrastive Preference Optimization (CPO) (Xu et al., 2024a), which has been used successfully to achieve very competitive translation performance. Our analysis seeks to describe the effects of preference-based fine-tuning on downstream performance, specifically regarding alignment effectiveness, the interactions between optimized and non-optimized metrics, and the impact of using multiple candidate translation systems for generating preference data. Through extensive experimentation, we find that:

- Preference-based alignment globally outperforms Supervised Fine-Tuning (SFT) on high-quality data in terms of maximizing the alignment metric.
- However, preference-based alignment is highly sensitive to the choice of candidate systems used for generating preference data, affecting both the alignment metric and downstream metric consistency.
- Aligning a model using its own translations achieves performance comparable to employing multiple external systems, while ensuring better metric consistency and allowing for improved control over the alignment process.

2 Background

2.1 Quality-Informed Translation

Along with human evaluation, lexical metrics like BLEU (Papineni et al., 2002), chrF (Popović, 2015), METEOR (Banerjee and Lavie, 2005), and ROUGE (Lin, 2004) have long been used for translation evaluation. However, human evaluation is costly, and lexical metrics have been shown to correlate poorly with human judgements.

More recently, some neural metrics have emerged as a preferred method to mimic human preferences without relying on expensive human evaluation. The intuitive approach involves training an encoder model on human-annotated source-translation-reference triplets. Among the metrics most frequently mentioned in the literature are BLEURT (Yan et al., 2023), COMET (Rei et al., 2020), CometKiwi (Rei et al., 2022b), xCOMET (Guerreiro et al., 2023), and Metric-X (Juraska et al., 2023). They can be divided into two families: *reference-based* metrics, that include a human-written gold reference as an input to the scoring model, and *reference-free* metrics, which only require access to the source sentence and the generated translation. These neural metrics have proven particularly effective at scoring translations and achieve much higher correlation with human judgments than their lexical counterparts (Mathur et al., 2020; Kocmi et al., 2021; Freitag et al., 2022b; Kocmi et al., 2024).

These neural metrics have also been leveraged to improve translation models through decoding strategies. The approach involves sampling various candidate translations, scoring them according to a given metric, and selecting the one with the highest score. This methodology is exemplified by MBR decoding in the reference-based setting and N-best reranking in the reference-free setting (Fernandes et al., 2022; Freitag et al., 2022a).

2.2 Quality-Based Fine-Tuning

With the recent rise of decoder-only LLMs applied to translation tasks (Zhu et al., 2023; Jiao et al., 2023; Hendy et al., 2023; Kocmi et al., 2023; Freitag et al., 2023; Xu et al., 2023; Alves et al., 2023; Xu et al., 2024a; Alves et al., 2024), and with automatic metrics increasingly reflecting human judgments (Sellam et al., 2020; Rei et al., 2020; Juraska et al., 2023), quality-based fine-tuning has gained considerable traction. This approach shifts the objective from selecting the best candidate translation according to a metric at inference time to directly updating model weights through fine-tuning to produce the desired translations. A straightforward approach is to perform SFT on high-quality translations, evaluated and then filtered with respect to a metric of interest (Alves et al., 2024).

Another attractive alternative is Preference Optimization (PO) (Simianer, 2018; Rafailov et al., 2024; Xu et al., 2024a; Yang et al., 2023; Xu et al., 2024b; Wu et al., 2024), which focuses on learning preferences between chosen and rejected translations rather than simply increasing the likelihood of high-quality sentences. A popular PO method is Direct Preference Optimization (DPO) (Rafailov et al., 2024), which aims to maximize a scaled likelihood gap between a chosen and a rejected option.

More recently, CPO (Xu et al., 2024a) has emerged as a promising alternative, incorporating an SFT term into the DPO loss, effectively combining the strengths of both methods. Moreover, by removing the reference policy from the learning objective, it improves training efficiency.

3 Experimental Setup

Here, we detail our experimental setup, explaining how we built the preference data, and train and evaluate the models.

3.1 Preference Data

Preference datasets. To build a preference dataset, one needs candidate translations, an evaluation metric m to score these translations, and a method to select chosen and rejected hypotheses. We denote a candidate dataset by

$$\mathcal{D} = \{(x_i, \mathcal{Y}_i)\}_{i=1}^{N},$$

where x_i denotes the source sentence and $\mathcal{Y}_i$ is a set of candidate translations. One can then derive a preference dataset,

$$\mathcal{D}_{pref} = \{(x_i, y_i^r, y_i^c)\}_{i=1}^{N},$$

where $y_i^c \in \mathcal{Y}_i$ (chosen hypothesis) is a translation preferred to $y_i^r \in \mathcal{Y}_i$ (rejected hypothesis) according to a metric m and a given selection method.

Multi-system approach. In the multi-system scenario, we follow the setting outlined by Xu et al. (2024a). Candidate translations are generated using three different systems, namely ALMA-13B-LoRA (the base model we aim to align, referred to as Base) (Xu et al., 2023), GPT-4 (OpenAI, 2023), and the human-written gold reference (referred to as Ref). Formally, for all data samples,

$$\mathcal{Y}_i^{multi} = \left\{y_i^{Ref}, y_i^{Base}, y_i^{GPT\text{-}4}\right\}.$$

Then, for each sample, the three translations are evaluated with regard to m. The one with highest (resp. lowest) score is selected as the chosen (resp. rejected) hypothesis. Formally,

$$y_i^c = \underset{y \in \mathcal{Y}_i^{multi}}{\arg\max}\, m(y) \;\wedge\; y_i^r = \underset{y \in \mathcal{Y}_i^{multi}}{\arg\min}\, m(y)$$

Mono-system approach. In the mono-system setting, we solely rely on the base model for candidate generation. For each source sentence, $K = 50$ candidates are top-p-sampled ($p = 0.6$) with a temperature $\tau = 0.9$,[2] and are then ranked based on evaluation metric m. For all samples, this results in a set of candidates

$$\mathcal{Y}_i^{mono} = \{y_i^1, \cdots, y_i^K\},$$

where $y_i^1 \preceq \cdots \preceq y_i^K$ are sorted in increasing quality order, with no loss of generality. Preference pairs are then derived to ensure that $y_i^r \preceq y_i^{Base} \preceq y_i^c$ holds for all samples. Further details on the construction of mono-system preference datasets are given in Section 5 and Appendix B.1.

Source dataset. We rely on the FLORES-200-based (Team et al., 2022) dataset used in Xu et al. (2024a) as a primary data source. It includes over 20000 translation pairs spanning six languages (English (`en`), Czech (`cs`), German (`de`), Icelandic (`is`), Russian (`ru`), and Chinese (`zh`)) and covering ten language directions, either into-English (`xx-en`) or out-of-English (`en-xx`).

Alignment metrics. In line with Xu et al. (2024a), we rely on reference-free neural metrics, namely xCOMET-QE-XXL (Guerreiro et al., 2023) (referred to as xCOMET-QE), and the WMT'23 version of CometKiwi-XXL (Rei et al., 2023) (denoted by CometKiwi), as well as on a reference-based lexical metric, chrF (Popović, 2015).

3.2 Training

Learning objective. We focus our diagnosis on CPO (Xu et al., 2024a), which combines a preference term with a likelihood term and achieves state-of-the-art performance in preference-based metric alignment for translation tasks. The empirical loss function is formally expressed as:

$$\mathcal{L}_{CPO} = -\frac{1}{N}\sum_{i=1}^{N}\left[\log\sigma\left(\beta\log\frac{\pi_\theta\left(y_i^c|x_i\right)}{\pi_\theta\left(y_i^r|x_i\right)}\right)\right] + \mathcal{L}_{SFT},$$

where $\mathcal{L}_{SFT} = -\frac{1}{N}\sum_{i=1}^{N}\left[\log\pi_\theta\left(y_i^c|x_i\right)\right]$ is the negative-log-likelihood loss applied to chosen translations, π_θ is the model to fine-tune, σ is the sigmoid function and β is a hyperparameter. In our experiments, CPO alignment is consistently compared to vanilla SFT on chosen translations.[3]

Training parameters. We replicate the exact same parameters as the ones outlined by Xu et al. (2024a). ALMA-13B-LoRA is LoRA fine-tuned

[2] These are the default parameters used in the ALMA paper (Xu et al., 2023, 2024a).

[3] All our models are trained using the code implementation provided by Xu et al. (2024a).

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	87.80	80.86	58.53	91.91	81.17	49.49
Preferences induced with **xCOMET-QE**						
SFT	● 89.13	81.49	59.82	● 92.38	81.67	50.28
CPO	● *89.95*	*81.89*	59.83	● *92.75*	*83.60*	*47.69*
Preferences induced with **CometKiwi**						
SFT	89.26	● 81.70	60.01	92.44	● 81.93	50.49
CPO	*89.82*	● *82.04*	*60.22*	*92.19*	● *83.64*	*48.11*
Preferences induced with **chrF**						
SFT	87.61	80.82	● 56.97	92.20	81.70	● 50.30
CPO	*78.51*	*75.62*	● *45.32*	*88.89*	*80.99*	● *42.50*

Table 1: Comparison between SFT on preferred translations and CPO in the multi-system setting, using xCOMET-QE, CometKiwi and chrF as alignment metrics. The same 3 metrics are reported for evaluation, separately for into-English (xx-en) and out-of-English (en-xx) translations on the WMT'22 dataset. Green shades indicate metric improvements over the base model, while red shades indicate metric decreases. We represent with (●) scenarios where the preference metric matches the evaluation metric. Values in *italic* font denote statistically significant differences between SFT- and CPO-based alignment at the 5% level, based on one-tailed paired Student's t-tests.

with rank 16 for one epoch, starting with a learning rate of 10^{-4}, using inverse square root decay and a batch size of 128. The β parameter of the CPO objective function is set equal to 0.1, in line with the original DPO paper by Rafailov et al. (2024).

3.3 Evaluation

Inference setup. Following other works on LLM-based translation (Alves et al., 2024; Briakou et al., 2024), all generations at inference time are produced using greedy decoding, as it provides maximum computational efficiency while preserving high output quality.[4]

Evaluation datasets. We evaluate our approaches on the WMT'22 test dataset, which consists of 17471 source-reference pairs and includes the same ten language pairs as the preference data. Evaluations on WMT'23 test data are provided in Appendix A.

Evaluation metrics. We use the same three metrics used to create the preference datasets: xCOMET-QE, CometKiwi, and chrF. Additional evaluation metrics are reported in Appendix A, specifically the reference-based version of Metric-X-Large (referred to as Metric-X) (Juraska et al., 2023), and BLEU (Papineni et al., 2002).

[4]Inference is performed using the vLLM library (Kwon et al., 2023).

4 Multi-System Preference Fine-Tuning

We begin our analysis by focusing on the multi-system setting (Xu et al., 2024a), in which the chosen and rejected options are derived from a pool of three candidate systems consisting of ALMA-13B-LoRA (base model), GPT-4, and the gold reference.

4.1 Top-Level Analysis

Neural-based alignment improves downstream performance. Table 1 shows that when aligning with neural metrics (xCOMET-QE or CometKiwi), both SFT on preferred translations and CPO consistently improve performance on the alignment metric across language pairs. We also observe that aligning on xCOMET-QE improves results on CometKiwi, and vice-versa. We hypothesize this may be the result of high correlation between different neural metrics, as they are typically trained on similar data. Overall, these results demonstrate that alignment-based techniques can achieve similar objectives to those of quality-aware decoding approaches like MBR, even though the target metric is only indirectly optimized.

CPO induces adverse metric effects. In Table 1, we observe that when aligning with neural metrics, CPO yields significantly greater improvements on the alignment metric compared to SFT. The inclusion of the reject option seems to offer additional benefits over the traditional SFT objective

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	87.80	80.86	58.53	91.91	81.17	49.49
Optimization via **SFT**						
Preferences induced with **xCOMET-QE**						
All systems	89.13	81.49	59.82	92.38	81.67	50.28
No Base	*89.41*	81.56	*60.26*	92.32	81.65	*50.52*
No Ref	*89.32*	*81.58*	*60.08*	*92.22*	*81.33*	*50.05*
No GPT-4	*88.44*	*81.15*	*58.86*	92.33	81.74	*50.06*
Preferences induced with **chrF**						
All systems	87.61	80.82	56.97	92.20	81.70	50.30
No Ref	*89.21*	*81.49*	*60.17*	91.99	*80.96*	*50.57*
Optimization via **CPO**						
Preferences induced with **xCOMET-QE**						
All systems	89.95	81.89	59.83	92.75	83.60	47.69
No Base	*89.59*	*81.73*	*59.94*	92.74	*83.13*	*48.54*
No Ref	89.91	81.86	*60.59*	*92.44*	*81.97*	*50.67*
No GPT-4	*88.81*	*81.35*	*57.91*	*92.22*	*83.16*	*46.82*
Preferences induced with **chrF**						
All systems	78.51	75.62	45.32	88.89	80.99	42.50
No Ref	*89.26*	*81.52*	*60.63*	*90.83*	*79.37*	*51.11*

Table 2: Impact of the systems used for candidate generation on WMT'22 performance in the multi-system setting after undergoing SFT and CPO optimization. Values in *italic* font denote statistically significant differences between all-systems-based alignment and alignment with one system removed, at the 5% significance level, based on one-tailed paired Student's t-tests. Evaluation metrics and color codes are the same as in Table 1.

in this context. However, aligning with CPO also introduces adverse effects between neural and lexical metrics for out-of-English translations. More specifically, and consistent with the findings of Xu et al. (2024a), aligning on neural metrics negatively impacts lexical metrics. Importantly, this is further evidence to support recommendations provided in (Kocmi et al., 2024): even though, in most cases, neural and lexical MT evaluation metrics should be positively correlated, we should employ caution when using the same metric for evaluation that was used during training/inference. Nevertheless and perhaps more interestingly, it turns out SFT does not produce such effects, raising the question of whether these contradictory evaluation dynamics seen with CPO stem from the learning objective itself or the mix of candidate systems used.

Lexical alignment fails to improve downstream performance. Table 1 shows that preference-based lexical alignment[5] behaves differently compared to neural alignment. Specifically, SFT results are roughly stagnant, showing a slight decrease in chrF for into-English translations and a slight increase for out-of-English translations. In contrast, CPO results in a steep drop across the metric board for both into- and out-of-English translations. Using the gold reference as the chosen system appears to impair downstream performance, especially when performing alignment using CPO.

4.2 Impact of the Candidate Systems

We now turn to investigating how much the success of alignment-based fine-tuning depends on the choice of the candidate systems. Unless otherwise specified, we use xCOMET-QE as the alignment metric and examine the performance impact of withdrawing systems from the candidate pool. We perform SFT and CPO on the newly created datasets. We report results in Table 2.

The choice of the candidate systems impacts alignment performance. Table 2 shows that for both SFT- and CPO-based methods, removing systems from the pool of candidates significantly af-

[5]When performing alignment using a lexical metric like chrF, the chosen translation is by definition the gold reference as long as it is present in the pool of candidates. The translation with the lowest chrF score among the remaining systems is then rejected.

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	87.80	80.86	58.53	91.91	81.17	49.49
Chosen system set to **Base**						
SFT	88.17	81.08	58.91	91.94	81.21	49.35
CPO	*87.94*	81.02	*58.62*	*91.75*	*81.06*	*48.56*
Chosen system set to **Ref**						
SFT	88.04	81.06	57.73	92.35	81.94	50.12
CPO	*81.95*	*77.86*	*48.75*	*86.97*	*80.01*	*39.81*
Chosen system set to **GPT-4**						
SFT	89.81	81.67	60.53	91.96	80.83	50.73
CPO	89.69	*80.99*	60.42	*90.50*	*78.81*	*50.22*

Table 3: Impact of imposing the chosen system on WMT'22 downstream performance in the multi-system setting. Values in *italic* font denote statistically significant differences between SFT- and CPO-based alignment at the 5% significance level, based on one-tailed paired Student's t-tests. Evaluation metrics and color codes are the same as in Table 1.

fects performance on the alignment metric. This is particularly the case for out-of-English translation with CPO optimization. Notably, removing GPT-4 has the strongest negative impact on downstream xCOMET-QE. This is expected as it is the highest-quality system among the system candidates (see Table 11 in Appendix B).

Some candidate systems can be harmful to preference-based alignment. In Section 4.1, we observed CPO negatively impacts en-xx chrF when aligning on neural metrics, unlike SFT on preferred translations. Table 2 suggests this may stem from including gold references in the candidate system pool: removing them eliminates this adverse effect. We also noted in Section 4.1 that lexical alignment fails to improve downstream chrF, with sharp decreases with CPO. This issue is resolved by removing gold references. Overall, candidate system choice affects alignment effectiveness and downstream metric consistency, with CPO showing higher sensitivity to preference settings than SFT.

4.3 Impact of the Chosen System

To complement findings from Section 4.2 and further characterize the sensitivity of preference-based alignment, we propose examining downstream performance when the chosen system is fixed to a single system. We create three preference datasets based on xCOMET-QE, in which we either impose the base model, reference or GPT-4 as the chosen system. When applicable, the rejected translation is selected from the remaining systems (if one has a lower xCOMET-QE than the chosen system); otherwise, the sample is discarded.

CPO is not robust to the preference setting. In contrast to the observations made in Section 4.1, Table 3 shows that, under this setup, CPO fails to outperform SFT for both xx-en and en-xx translations. When systematically choosing base translations, CPO is unable to surpass the trivial SFT setting where the base model is fine-tuned on its own translations.[6] Moreove, downstream CPO performance significantly declines when gold references are chosen, underperforming the non-aligned model across all metrics, even including the alignment metric. These results reinforce the claims made in Section 4.2 and indicate a lack of robustness of CPO compared to SFT. In the following section (Section 5), we demonstrate that this instability observed with CPO can be mitigated by using a more normalized preference setting, relying only on the base model for candidate generation.

5 Mono-System Preference Fine-Tuning

So far, we have exclusively focused on multi-system alignment, which involves using external models for candidate generation and preference dataset building. Although this approach is common for metric alignment (Luong and Manning,

[6] As expected, performing SFT on a model's own greedy predictions has minimal impact on downstream performance.

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	87.80	80.86	58.53	91.91	81.17	49.49
Optimization via **SFT**						
Multi-system	89.13	81.49	59.82	92.38	81.67	50.28
Mono-system	*88.51*	*81.29*	*59.05*	*92.17*	81.54	*49.41*
Optimization via **CPO**						
Multi-system	89.95	81.89	59.83	92.75	83.60	47.69
Mono-system	*89.35*	81.80	*59.52*	92.69	*82.91*	*49.02*
Mono-system (opt.)	*89.58*	81.97	*59.65*	92.87	83.47	*49.11*

Table 4: Comparison between multi- and mono-system fine-tuning on WMT'22 test data. Alignment is performed on xCOMET-QE for both SFT and CPO. Mono-system (opt.) denotes the model fine-tuned on optimized mono-system preference data. Values in *italic* font denote statistically significant differences between multi-system- and mono-system-based alignment at the 5% significance level. Evaluation metrics and color codes are the same as in Table 1, based on one-tailed paired Student's t-tests.

2015; Sennrich et al., 2016; Xu et al., 2024a), some works have shown that a model can be aligned effectively using only its own outputs (Yang et al., 2023; Yuan et al., 2024; Dubey et al., 2024). In this section, we propose to take a closer look at this strategy and identify its potential advantages and disadvantages compared to the multi-system approach. We use xCOMET-QE as the alignment metric. To ensure a fair comparison, we first generate the mono-system dataset to approximately replicate the properties of the multi-system dataset regarding the alignment metric.[7] Details on the construction of mono-system preference datasets are given in Section 3 and Appendix B.1.

5.1 Comparison With Multi-System Alignment

Mono-system alignment improves downstream performance. Table 4 shows that performing SFT and CPO on a mono-system dataset using xCOMET-QE for alignment results in improved downstream performance across all neural metrics compared to the base model, as observed in the multi-system scenario (Section 4.1). This finding highlights the effectiveness of alignment techniques even when using only the model's own translations for candidate generation, without needing access to high-quality external systems. This is particularly relevant in practical scenarios in which such access may be limited or unavailable.

[7]The created mono-system dataset has an average rejected/-chosen xCOMET-QE of 87.8/97.3, compared to 87.9/97.2 for the multi-system dataset (Table 11).

CPO consistently outperforms SFT on neural metrics. Similar to when relying on multiple systems for candidate generation, we observe in Table 4 that CPO outperforms SFT regarding downstream performance on neural metrics. This finding reinforces the observation made in Section 4.1 and tends to confirm the superiority of the CPO objective over SFT on preferred translations in optimizing neural-based alignment performance.

Mono-system alignment slightly underperforms multi-system alignment. Table 4 shows that while mono-system alignment increases downstream performance on neural metrics, the improvement levels are not as high as in the multi-system setting. Despite the mono- and multi-system preference datasets being built with the same alignment metric properties, having translations from different distributions, particularly from GPT-4 (cf. Section 4.2 and Table 2), appears to add value for achieving optimized alignment effectiveness.

Removing external systems almost eliminates the adverse metric effects observed with CPO. In Section 4.1, we showed that multi-system neural alignment using CPO greatly impacts lexical performance for out-of-English translations. Table 4 demonstrates that mono-system alignment almost completely mitigates these negative effects. While there is still a slight decrease in en-xx chrF, it is much smaller compared to the multi-system scenario. This confirms the findings from Sections 4.2 and 4.3 that CPO is sensitive to the preference setting, but also shows that relying solely on candidate translations from the base model limits adverse ef-

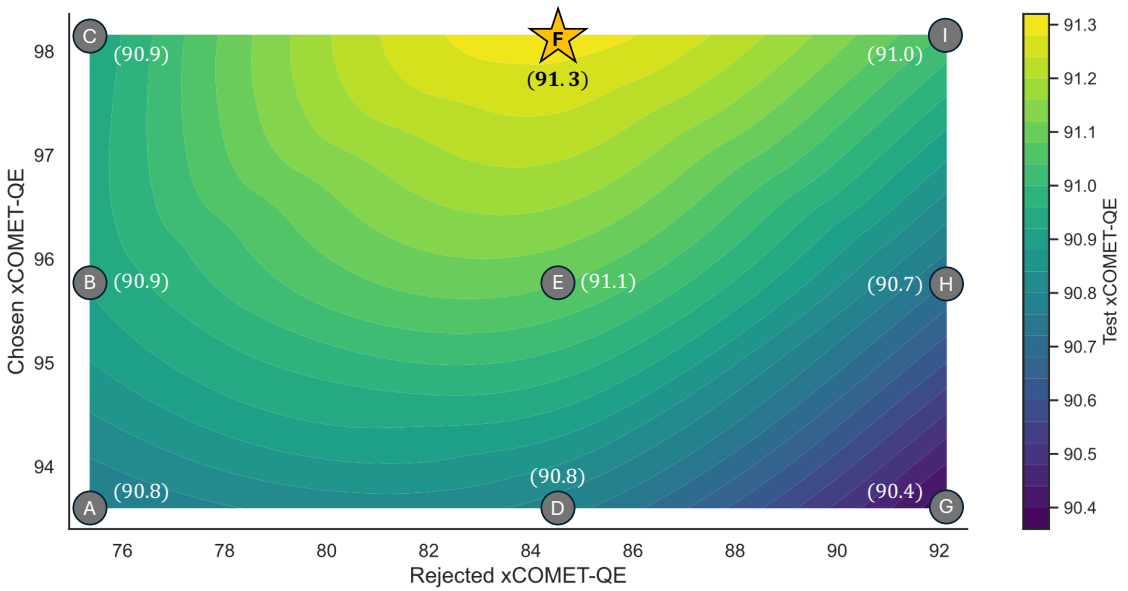

Figure 1: Impact of chosen and rejected option quality on downstream performance, using xCOMET-QE for alignment and evaluation. The chart is derived by linearly interpolating results from nine preference datasets (points A to I), each with different average rejected and chosen qualities. Test performance on WMT'22 (average across all language pairs) is reported in brackets. Example: point C (avg. rejected xCOMET-QE: 75.4, avg. chosen: 98.2) achieves 90.9 xCOMET-QE on WMT'22 test data.

fects on downstream metric consistency. A possible explanation is that candidate translations from the same system distribution tend to have similar properties, thereby reducing the likelihood of observing high lexical instability when performing alignment based on a neural metric like xCOMET-QE.

The mono-system approach offers better control over the alignment process. Specifically, mono-system alignment provides more fine-grained control over the respective qualities of the chosen and rejected options. This setting allows for tuning these qualities to maximize post-alignment performance, which is not possible when using a limited number of external systems. This aspect is further explored in the following section (Section 5.2).

5.2 Optimizing the Preference Data

In this final experiment, we examine how the quality of chosen and rejected options affects downstream performance. We build nine preference datasets, each with varying average xCOMET-QE scores for chosen and rejected options. The hypotheses' average qualities are categorized into three groups: High, Mid, and Low. As detailed in Section 3.1, the quality of the chosen (resp. rejected) option is always ensured to be above (resp. below) the quality of the base translation. The statistics of the created datasets are summarized in Appendix B.1 (Table 11).

The respective qualities of the rejected and chosen options have a significant impact on post-CPO performance. Figure 1 highlights the need to closely monitor the qualities of chosen and rejected options to fully leverage the mono-system approach. Specifically, several properties of preference data were found to negatively impact post-CPO performance: (i) a chosen option of too low quality, (ii) an extremely low or high quality of the rejected option, and (iii) too wide a gap between the qualities of the rejected and chosen options.

Optimizing preference data yields competitive performance to multi-system setting. Figure 1 shows that for effective metric alignment with CPO, the rejected option's quality should be moderate (neither too high nor too low), while the chosen option's quality should be as high as possible. Specifically, optimal test performance was obtained with rejected options average around 90% ($\Delta = -10\%$) of the base model's quality, and chosen options averaging around 105% ($\Delta = +5\%$). Under this scenario, we show that performance levels can match those in the multi-system setting while maintaining consistency with lexical scores (Table 4). However, these results also highlight the complexity of achieving optimal preference-based alignment and get the most of the reject option.

6 Conclusion

Our experiments revealed several key findings. Firstly, we showed that preference-based alignment, specifically using CPO, globally outperforms SFT on high-quality data in terms of improving neural evaluation metrics. However, we identified significant drawbacks when relying on multiple systems for preference data generation, revealing adverse effects between neural and lexical metrics, and highlighting a lack of robustness in preference-based alignment compared to the SFT approach. Finally, we showed that using candidate translations all originating from the same system distribution, specifically the base model, can be an effective strategy for gaining more control over preference-based fine-tuning. This approach achieves performance comparable to using multiple external systems while ensuring better consistency across evaluation metrics. In a nutshell, while preference-based alignment techniques hold promise for improving MT quality, careful consideration must be given to the choice of candidate translations, the learning objective, and the potential trade-offs regarding downstream metric consistency.

Limitations

In this work, we conducted extensive experiments to assess the impact of preference-based fine-tuning on downstream translation quality. For efficiency and practicality, we focused on the experimental setup detailed by Xu et al. (2024a), which utilizes three systems for candidate generation. Similarly, we used the same evaluation metrics and datasets. Future experiments could benefit from validating our findings using different model families, a broader range of alignment and evaluation metrics, and additional translation datasets, for instance including other languages.

Additionally, in the mono-system setting, we explored the impact of varying the qualities of chosen and rejected options and derived general insights on optimizing preference data. Further research could involve using different datasets, models, and alignment metrics to characterize more precisely the factors that influence downstream performance in this specific scenario. This approach could lead to a deeper mathematical understanding of the elements that affect performance in preference-based fine-tuning, resulting in more robust and scalable optimization techniques.

Finally, our evaluation relied on automatic metrics, both lexical and neural, with the latter closely approximating human judgments but still being unable to fully replace them. Given their imperfect correlation with human preferences, future work could benefit from additional human evaluation of outputs obtained via the approaches we studied to get an even deeper understanding of post-alignment downstream performance dynamics.

Ethics Statement

Our work aims to investigate the mechanisms of model alignment to enhance transparency in the field of automatic translation. We believe this effort improves the interpretability of model outputs, which is beneficial for ethical considerations. Additionally, our analysis is distinctly multilingual, with an emphasis on low-resource languages, contributing to expanding the scope of MT. We have identified no potential negative societal impacts from our work.

Acknowledgements

Training compute was provided by the Jean Zay supercomputer, operated by GENCI IDRIS, through compute grants 2023-AD011014668R1 and AD010614770, as well as by the Adastra supercomputer through projects c1615122, cad15031, and cad14770. Part of this work was also supported by the EU's Horizon Europe Research and Innovation Actions (UTTER, contract 101070631), by the DECOLLAGE project (ERC-2022-CoG 101088763), and by the Portuguese Recovery and Resilience Plan through project C645008882-00000055 (Center for Responsible AI).

References

Duarte M Alves, Nuno M Guerreiro, João Alves, José Pombal, Ricardo Rei, José GC de Souza, Pierre Colombo, and André FT Martins. 2023. Steering large language models for machine translation with finetuning and in-context learning. *arXiv preprint arXiv:2310.13448*.

Duarte M Alves, José Pombal, Nuno M Guerreiro, Pedro H Martins, João Alves, Amin Farajian, Ben Peters, Ricardo Rei, Patrick Fernandes, Sweta Agrawal, et al. 2024. Tower: An open multilingual large language model for translation-related tasks. *arXiv preprint arXiv:2402.17733*.

Chantal Amrhein and Rico Sennrich. 2022. Identifying weaknesses in machine translation metrics through minimum bayes risk decoding: A case study for comet. *arXiv preprint arXiv:2202.05148*.

Satanjeev Banerjee and Alon Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.

Eleftheria Briakou, Jiaming Luo, Colin Cherry, and Markus Freitag. 2024. Translating step-by-step: Decomposing the translation process for improved translation quality of long-form texts. *arXiv preprint arXiv:2409.06790*.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaoqing Ellen Tan, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aaron Grattafiori, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alex Vaughan, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Franco, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, Danny Wyatt, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Firat Ozgenel, Francesco Caggioni, Francisco Guzmán, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Govind Thattai, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou

U, Karan Saxena, Karthik Prasad, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kun Huang, Kunal Chawla, Kushal Lakhotia, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Maria Tsimpoukelli, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikolay Pavlovich Laptev, Ning Dong, Ning Zhang, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Rohan Maheswari, Russ Howes, Ruty Rinott, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Kohler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vítor Albiero, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaofang Wang, Xiaojian Wu, Xiaolan Wang, Xide Xia, Xilun Wu, Xinbo Gao, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yuchen Hao, Yundi Qian, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, and Zhiwei Zhao. 2024. The llama 3 herd of models.

Bryan Eikema and Wilker Aziz. 2020. Is MAP decoding all you need? the inadequacy of the mode in neural machine translation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4506–4520, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Patrick Fernandes, António Farinhas, Ricardo Rei, José G. C. de Souza, Perez Ogayo, Graham Neubig, and Andre Martins. 2022. Quality-aware decoding for neural machine translation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1396–1412, Seattle, United States. Association for Computational Linguistics.

Mara Finkelstein, Subhajit Naskar, Mehdi Mirzazadeh, Apurva Shah, and Markus Freitag. 2024. Mbr and qe finetuning: Training-time distillation of the best and most expensive decoding methods.

Markus Freitag, David Grangier, Qijun Tan, and Bowen Liang. 2022a. High quality rather than high model probability: Minimum Bayes risk decoding with neural metrics. *Transactions of the Association for Computational Linguistics*, 10:811–825.

Markus Freitag, Nitika Mathur, Chi-kiu Lo, Eleftherios Avramidis, Ricardo Rei, Brian Thompson, Tom Kocmi, Frederic Blain, Daniel Deutsch, Craig Stewart, Chrysoula Zerva, Sheila Castilho, Alon Lavie, and George Foster. 2023. Results of WMT23 metrics shared task: Metrics might be guilty but references are not innocent. In *Proceedings of the Eighth Conference on Machine Translation*, pages 578–628, Singapore. Association for Computational Linguistics.

Markus Freitag, Ricardo Rei, Nitika Mathur, Chi-kiu Lo, Craig Stewart, Eleftherios Avramidis, Tom Kocmi, George Foster, Alon Lavie, and André F. T. Martins. 2022b. Results of WMT22 metrics shared task: Stop using BLEU – neural metrics are better and more robust. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 46–68, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Nuno M Guerreiro, Ricardo Rei, Daan van Stigt, Luisa Coheur, Pierre Colombo, and André FT Martins. 2023. xcomet: Transparent machine translation evaluation through fine-grained error detection. *arXiv preprint arXiv:2310.10482*.

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

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

Juraj Juraska, Mara Finkelstein, Daniel Deutsch, Aditya Siddhant, Mehdi Mirzazadeh, and Markus Freitag. 2023. MetricX-23: The Google submission to the WMT 2023 metrics shared task. In *Proceedings of the Eighth Conference on Machine Translation*, pages 756–767, Singapore. Association for Computational Linguistics.

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Philipp Koehn, Benjamin Marie, Christof Monz, Makoto Morishita, Kenton Murray, Makoto Nagata, Toshiaki Nakazawa, Martin Popel, Maja Popović, and Mariya Shmatova. 2023. Findings of the 2023 conference on machine translation (WMT23): LLMs are here but not quite there yet. In *Proceedings of the Eighth Conference on Machine Translation*, pages 1–42, Singapore. Association for Computational Linguistics.

Tom Kocmi, Christian Federmann, Roman Grundkiewicz, Marcin Junczys-Dowmunt, Hitokazu Matsushita, and Arul Menezes. 2021. To ship or not to ship: An extensive evaluation of automatic metrics for machine translation. In *Proceedings of the Sixth Conference on Machine Translation*, pages 478–494, Online. Association for Computational Linguistics.

Tom Kocmi, Vilém Zouhar, Christian Federmann, and Matt Post. 2024. Navigating the metrics maze: Reconciling score magnitudes and accuracies. *arXiv preprint arXiv:2401.06760*.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39, Vancouver. Association for Computational Linguistics.

Shankar Kumar and Bill Byrne. 2002. Minimum bayes-risk word alignments of bilingual texts. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 140–147.

Shankar Kumar and William Byrne. 2004. Minimum Bayes-risk decoding for statistical machine translation. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 169–176, Boston, Massachusetts, USA. Association for Computational Linguistics.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Minh-Thang Luong and Christopher Manning. 2015. Stanford neural machine translation systems for spoken language domains. In *Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign*, pages 76–79, Da Nang, Vietnam.

Nitika Mathur, Timothy Baldwin, and Trevor Cohn. 2020. Tangled up in BLEU: Reevaluating the evaluation of automatic machine translation evaluation metrics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4984–4997, Online. Association for Computational Linguistics.

R OpenAI. 2023. Gpt-4 technical report. arxiv 2303.08774. *View in Article*, 2(5).

Myle Ott, Michael Auli, David Grangier, and Marc'Aurelio Ranzato. 2018. Analyzing uncertainty in neural machine translation. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3956–3965. PMLR.

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.

Maja Popović. 2015. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the tenth workshop on statistical machine translation*, pages 392–395.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Ricardo Rei, José G. C. de Souza, Duarte Alves, Chrysoula Zerva, Ana C Farinha, Taisiya Glushkova, Alon Lavie, Luisa Coheur, and André F. T. Martins. 2022a. COMET-22: Unbabel-IST 2022 submission for the metrics shared task. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 578–585, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Ricardo Rei, Nuno M. Guerreiro, JosÃ© Pombal, Daan van Stigt, Marcos Treviso, Luisa Coheur, José G. C. de Souza, and André Martins. 2023. Scaling up CometKiwi: Unbabel-IST 2023 submission for the quality estimation shared task. In *Proceedings of the Eighth Conference on Machine Translation*, pages 841–848, Singapore. Association for Computational Linguistics.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Ricardo Rei, Marcos Treviso, Nuno M. Guerreiro, Chrysoula Zerva, Ana C Farinha, Christine Maroti, José G. C. de Souza, Taisiya Glushkova, Duarte Alves, Luisa Coheur, Alon Lavie, and André F. T. Martins. 2022b. CometKiwi: IST-unbabel 2022 submission for the quality estimation shared task. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 634–645, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Thibault Sellam, Dipanjan Das, and Ankur Parikh. 2020. BLEURT: Learning robust metrics for text generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7881–7892, Online. Association for Computational Linguistics.

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.

Patrick Simianer. 2018. *Preference Learning for Machine Translation*. Ph.D. thesis.

NLLB Team, Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, et al. 2022. No language left behind: Scaling human-centered machine translation (2022). *URL https://arxiv. org/abs/2207.04672*.

Qiyu Wu, Masaaki Nagata, Zhongtao Miao, and Yoshimasa Tsuruoka. 2024. Word alignment as preference for machine translation. *arXiv preprint arXiv:2405.09223*.

Haoran Xu, Young Jin Kim, Amr Sharaf, and Hany Hassan Awadalla. 2023. A paradigm shift in machine translation: Boosting translation performance of large language models. *arXiv preprint arXiv:2309.11674*.

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

Nuo Xu, Jun Zhao, Can Zu, Tao Gui, Qi Zhang, and Xuanjing Huang. 2024b. Advancing translation preference modeling with rlhf: A step towards cost-effective solution. *arXiv preprint arXiv:2402.11525*.

Yiming Yan, Tao Wang, Chengqi Zhao, Shujian Huang, Jiajun Chen, and Mingxuan Wang. 2023. BLEURT has universal translations: An analysis of automatic metrics by minimum risk training. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5428–5443, Toronto, Canada. Association for Computational Linguistics.

Guangyu Yang, Jinghong Chen, Weizhe Lin, and Bill Byrne. 2023. Direct preference optimization for neural machine translation with minimum bayes risk decoding. *arXiv preprint arXiv:2311.08380*.

Weizhe Yuan, Richard Yuanzhe Pang, Kyunghyun Cho, Xian Li, Sainbayar Sukhbaatar, Jing Xu, and Jason Weston. 2024. Self-rewarding language models.

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

A Additional Results

In this section, we present results on WMT'23 test data. The findings in Tables 5, 6, 7 and 8 support the observations discussed in the main text for the WMT'22 dataset. In Tables 9 and 10, we also provide additional insights, split by language pairs, and include extra metrics, specifically Metric-X and BLEU.

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	88.00	77.74	52.30	86.19	73.08	47.31
Preferences induced with **xCOMET-QE**						
SFT	• 88.96	78.46	53.30	• 87.07	73.99	48.38
CPO	• *89.77*	*78.95*	53.47	• *88.09*	*76.75*	*44.29*
Preferences induced with **CometKiwi**						
SFT	89.03	• 78.57	53.53	87.11	• 74.21	48.45
CPO	*89.58*	• *79.16*	*53.97*	87.25	• *76.71*	*44.48*
Preferences induced with **chrF**						
SFT	87.91	77.62	• 51.20	86.95	73.96	• 48.14
CPO	*81.79*	*72.38*	• *41.46*	*83.21*	*74.76*	• *37.96*

Table 5: Comparison between SFT on preferred translations and CPO in the multi-system setting on WMT'23 test data. Notations and formatting are the same as in Table 1.

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	88.00	77.74	52.30	86.19	73.08	47.31
Optimization via **SFT**						
Preferences induced with **xCOMET-QE**						
All systems	• 88.96	78.46	53.30	• 87.07	73.99	48.38
No Base	• 89.07	78.53	*53.57*	• *86.94*	*73.70*	*48.52*
No-Ref	• 89.05	78.47	53.39	• 87.04	*73.60*	*48.65*
No GPT-4	• *88.29*	*78.02*	*52.62*	• 87.04	74.08	*48.03*
Preferences induced with **chrF**						
All systems	87.91	77.62	• 51.20	86.95	73.96	• 48.14
No Ref	*88.89*	*78.47*	• *53.51*	86.65	*73.02*	• *49.04*
Optimization via **CPO**						
Preferences induced with **xCOMET-QE**						
All systems	• 89.77	78.95	53.47	• 88.09	76.75	44.29
No Base	• *89.52*	*78.54*	53.44	• *87.66*	*75.84*	*45.27*
No Ref	• *89.57*	*79.26*	*54.18*	• *87.41*	*74.46*	*48.88*
No GPT-4	• *89.16*	*78.46*	*51.94*	• *87.45*	76.62	*43.30*
Preferences induced with **chrF**						
All systems	81.79	72.38	• 41.46	83.21	74.76	• 37.96
No Ref	*88.79*	*78.73*	• *54.21*	*85.40*	71.82	• *49.59*

Table 6: Impact of candidate systems on WMT'23 downstream performance in the multi-system setting. Notations and formatting are the same as in Table 2.

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	88.00	77.74	52.30	86.19	73.08	47.31
Chosen system set to **Base**						
SFT	• 88.07	77.93	52.52	• 86.52	73.27	47.52
CPO	• 88.05	77.95	*52.24*	• *86.68*	*73.75*	*46.54*
Chosen system set to **Ref**						
SFT	• 88.33	77.92	51.75	• 87.29	74.57	47.86
CPO	• *84.06*	74.22	*44.53*	• *81.01*	*73.55*	*34.64*
Chosen system set to **GPT-4**						
SFT	• 89.57	79.06	54.08	• 86.70	73.18	49.23
CPO	• *88.99*	*78.64*	53.95	• *85.14*	*71.40*	*48.68*

Table 7: Impact of the chosen system on WMT'23 downstream performance in the multi-system setting. Notations and formatting are the same as in Table 3.

	xx-en			en-xx		
	Neural		Lexical	Neural		Lexical
	xCOMET-QE	CometKiwi	chrF	xCOMET-QE	CometKiwi	chrF
Base	88.00	77.74	52.30	86.19	73.08	47.31
Optimization via **SFT**						
Multi-system	• 88.96	78.46	53.30	• 87.07	73.99	48.38
Mono-system	• *88.55*	*78.17*	*52.74*	• *86.75*	73.87	*47.43*
Optimization via **CPO**						
Multi-system	• 89.77	78.95	53.47	• 88.09	76.75	44.29
Mono-system	• *89.33*	*78.78*	*53.17*	• 87.94	*76.01*	*46.65*
Mono-system (opt.)	• *89.36*	78.92	53.28	• *88.50*	76.87	*46.48*

Table 8: Comparison between multi- and mono-system fine-tuning on WMT'23 test data. Notations and formatting are the same as in Table 4.

	cs-en					en-cs					de-en					en-de				
	Neural			Lexical		Neural			Lexical		Neural			Lexical		Neural			Lexical	
	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU
Base	83.42	82.57	2.00	65.81	41.25	90.80	81.96	1.48	53.58	27.42	93.33	83.48	2.03	55.24	29.02	96.32	80.87	1.22	56.95	27.65
SFT																				
Multi-system																				
→ xCOMET-QE																				
Vanilla	86.18	83.17	1.98	67.36	42.81	91.49	82.69	1.42	55.02	28.58	93.88	83.85	2.00	56.32	29.86	96.57	81.21	1.19	57.62	28.22
No Base	86.58	83.01	1.99	67.79	42.99	91.61	82.63	1.42	55.51	28.84	93.88	83.88	2.00	56.62	30.08	96.63	81.66	1.19	57.83	28.15
No Ref	86.43	83.20	1.97	67.50	42.93	91.28	82.10	1.45	54.94	28.75	93.88	83.96	1.99	56.63	30.06	96.41	81.10	1.20	57.49	28.08
No GPT-4	85.23	83.03	2.02	66.55	42.21	91.53	82.78	1.42	54.52	28.07	93.41	83.59	2.02	55.41	29.10	96.49	81.22	1.20	57.38	28.14
Chosen = Base	84.64	82.87	2.03	66.39	41.99	90.77	81.69	1.50	53.52	27.67	93.55	83.61	2.03	55.82	29.48	96.43	81.14	1.22	56.97	27.80
Chosen = Ref	83.78	82.83	2.04	65.74	41.62	91.36	82.65	1.40	54.35	27.73	93.35	83.69	2.02	54.67	28.48	96.67	81.60	1.19	57.32	27.99
Chosen = GPT-4	86.97	82.69	2.00	67.62	41.99	91.14	81.33	1.48	56.27	28.97	94.22	83.95	1.98	57.02	30.10	96.29	80.92	1.21	58.05	28.00
→ CometKiwi																				
Vanilla	86.17	83.31	1.97	67.50	43.08	91.57	83.01	1.40	55.29	28.69	94.00	84.07	1.98	56.40	29.79	96.58	81.47	1.19	57.77	28.13
→ chrF																				
Vanilla	83.63	82.85	2.06	65.22	40.77	91.42	82.45	1.41	54.76	28.11	92.86	83.53	2.03	54.26	28.29	96.60	81.63	1.18	57.50	28.06
No Ref	86.15	82.91	1.99	67.67	43.23	91.29	81.85	1.48	55.83	29.24	93.91	83.85	2.00	56.58	30.04	96.32	80.74	1.22	57.97	28.21
Mono-system																				
→ xCOMET-QE																				
Vanilla	85.00	83.11	1.99	66.38	41.81	91.04	82.23	1.45	53.27	27.42	93.71	83.80	2.00	55.84	29.27	96.46	81.16	1.21	57.08	27.94
CPO																				
Multi-system																				
→ xCOMET-QE																				
Vanilla	87.40	83.58	1.94	67.52	42.54	90.86	84.58	1.40	50.58	23.44	94.22	84.10	1.94	56.21	29.44	97.34	83.38	1.12	55.95	26.20
No Ref	87.59	83.18	1.95	67.90	41.86	91.60	82.63	1.41	56.16	28.97	94.28	84.07	1.98	56.98	30.02	96.57	82.02	1.18	57.86	27.66
No Base	86.68	83.22	1.96	67.46	42.84	90.94	84.14	1.38	51.48	24.91	93.97	84.06	1.96	56.32	29.80	97.16	82.67	1.12	56.53	27.20
No GPT-4	84.95	83.22	1.99	65.99	41.31	90.45	83.91	1.39	49.32	22.04	93.65	83.88	1.96	54.95	28.36	97.31	83.27	1.12	55.30	25.17
Chosen = Base	84.10	82.69	2.05	65.94	41.77	90.42	81.80	1.51	52.29	26.71	93.41	83.48	2.04	55.53	29.34	96.37	80.82	1.23	56.17	27.10
Chosen = Ref	71.82	79.56	2.22	54.77	28.59	79.03	80.03	1.59	40.19	14.25	89.61	81.47	2.08	47.32	21.40	96.57	79.91	1.12	48.62	18.46
Chosen = GPT-4	87.70	81.77	2.07	66.72	39.42	89.73	78.83	1.57	56.21	27.28	94.17	83.06	2.03	57.19	29.47	95.68	79.61	1.27	57.68	26.23
→ CometKiwi																				
Vanilla	86.74	83.46	1.95	67.67	42.42	90.21	84.81	1.43	51.23	23.53	94.20	84.22	1.93	56.45	29.61	97.19	83.71	1.13	56.24	25.41
→ chrF																				
Vanilla	66.31	76.41	2.53	51.03	25.35	83.47	80.25	1.72	43.35	15.80	87.27	79.30	2.29	44.28	19.34	96.87	81.49	1.14	51.01	19.73
No Ref	86.65	82.99	1.98	68.33	43.19	89.83	79.38	1.59	57.08	28.51	93.94	83.67	2.02	56.68	29.76	95.84	80.01	1.27	58.68	27.66
Mono-system																				
→ xCOMET-QE																				
Vanilla	85.99	83.66	1.93	67.10	42.09	91.40	83.99	1.39	52.46	26.30	94.01	84.14	1.95	55.97	29.18	96.89	82.29	1.15	56.53	27.25
Optimized	86.51	83.84	1.90	67.08	41.73	91.47	84.28	1.37	52.40	25.92	94.26	84.32	1.94	56.15	29.26	97.13	82.78	1.12	56.72	27.09

	is-en					en-is					ru-en					en-ru				
	Neural			Lexical		Neural			Lexical		Neural			Lexical		Neural			Lexical	
	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU
Base	76.22	85.36	1.89	59.72	35.34	89.15	80.68	2.40	53.31	23.49	89.68	80.92	1.82	62.72	35.29	92.77	82.62	2.04	51.93	25.86
SFT																				
Multi-system																				
→ xCOMET-QE																				
Vanilla	79.24	85.88	1.82	62.05	37.49	89.18	80.63	2.38	53.14	22.90	90.59	81.35	1.77	64.18	36.97	93.18	83.18	1.96	53.02	26.79
No Base	80.12	86.09	1.84	62.77	38.11	88.65	80.19	2.45	53.23	22.95	90.85	81.40	1.77	64.60	37.20	93.02	83.10	1.97	53.26	26.91
No Ref	79.67	85.94	1.83	62.52	37.97	88.74	80.10	2.47	52.93	22.84	90.82	81.50	1.77	64.47	37.24	92.90	82.73	2.04	52.57	26.29
No GPT-4	77.64	85.63	1.87	60.33	35.79	89.43	81.05	2.34	53.38	23.27	90.04	81.15	1.80	63.20	36.19	93.13	83.05	1.95	52.82	26.65
Chosen = Base	76.85	85.60	1.87	59.94	35.60	88.67	80.35	2.42	52.58	22.53	89.87	80.99	1.82	63.08	36.01	92.87	82.64	2.05	52.13	26.22
Chosen = Ref	76.86	85.52	1.87	59.88	35.41	89.44	81.24	2.30	53.45	23.60	89.39	80.92	1.80	61.40	34.43	93.29	83.43	1.92	52.74	26.54
Chosen = GPT-4	81.03	86.01	1.84	63.09	37.61	88.26	79.01	2.56	53.17	22.44	91.11	81.45	1.77	64.91	37.39	92.47	82.44	2.07	53.53	26.96
→ CometKiwi																				
Vanilla	79.38	86.03	1.81	62.26	37.69	89.06	80.95	2.37	53.42	23.30	90.78	81.52	1.75	64.31	37.09	93.20	83.30	1.96	53.22	26.96
→ chrF																				
Vanilla	76.29	85.36	1.89	59.07	34.61	88.99	80.46	2.35	53.09	23.20	88.63	80.51	1.84	60.40	33.44	93.11	83.31	1.93	52.80	26.77
No Ref	79.56	85.96	1.83	62.19	37.70	88.34	79.65	2.54	53.22	22.88	90.79	81.40	1.77	64.56	37.30	92.67	82.50	2.06	53.29	26.90
Mono-system																				
→ xCOMET-QE																				
Vanilla	77.44	85.77	1.85	60.19	35.69	89.11	80.46	2.38	52.95	23.16	90.25	81.27	1.78	63.52	36.22	93.14	83.07	2.00	52.13	26.40
CPO																				
Multi-system																				
→ xCOMET-QE																				
Vanilla	80.73	86.12	1.81	63.01	38.10	89.28	82.70	2.08	51.70	21.29	91.14	81.57	1.72	63.88	35.78	94.52	85.51	1.71	50.60	23.97
No Ref	81.03	85.91	1.84	63.29	37.90	89.15	80.18	2.47	53.47	22.46	91.11	81.61	1.77	64.66	36.55	92.89	83.40	1.95	53.27	26.27
No Base	79.90	86.07	1.80	62.63	37.81	89.67	82.02	2.10	51.95	22.34	91.06	81.57	1.73	64.20	36.49	94.30	84.85	1.81	51.28	25.44
No GPT-4	77.75	85.63	1.84	60.46	35.73	88.70	82.37	2.08	51.02	20.54	90.04	81.07	1.78	61.58	33.91	94.44	85.41	1.70	49.69	23.09
Chosen = Base	75.89	85.47	1.93	59.27	35.04	88.83	80.38	2.37	52.21	22.51	89.76	81.01	1.81	62.86	35.46	92.57	82.11	2.13	51.33	25.64
Chosen = Ref	62.78	82.24	2.09	51.44	25.64	76.02	77.65	2.46	43.95	13.38	83.04	77.41	2.01	51.27	23.21	93.39	82.63	1.73	42.24	15.76
Chosen = GPT-4	81.02	85.15	1.92	62.85	36.17	86.51	76.73	2.88	52.54	20.88	90.95	80.91	1.83	64.68	36.06	91.13	80.75	2.21	53.43	25.58
→ CometKiwi																				
Vanilla	80.83	86.17	1.79	63.15	37.98	88.15	82.29	2.30	52.23	21.16	91.06	81.80	1.72	64.15	36.14	93.99	85.44	1.77	51.04	23.54
→ chrF																				
Vanilla	57.08	80.43	2.40	47.93	22.21	81.59	78.80	2.34	46.13	14.70	79.00	75.29	2.25	47.38	20.28	93.52	83.83	1.82	45.51	17.75
No Ref	80.13	85.77	1.85	62.87	37.63	86.51	77.57	2.87	53.35	22.12	90.89	81.43	1.79	64.57	36.80	91.54	81.04	2.18	53.86	26.18
Mono-system																				
→ xCOMET-QE																				
Vanilla	79.33	86.05	1.80	61.78	37.42	89.75	82.17	2.14	52.71	22.69	90.78	81.44	1.75	63.50	35.72	93.79	83.99	1.87	52.03	26.09
Optimized	79.16	86.02	1.78	62.25	37.42	89.81	82.81	2.11	52.79	22.60	90.96	81.60	1.72	63.69	35.57	93.95	84.77	1.84	52.18	26.00

	zh-en					en-zh					xx-en					en-xx				
	Neural			Lexical		Neural			Lexical		Neural			Lexical		Neural			Lexical	
	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU
Base	89.49	74.32	4.10	51.25	21.69	89.10	79.48	2.30	33.62	34.52	87.80	80.86	2.42	58.53	31.78	91.91	81.17	1.83	49.49	28.28
SFT																				
Multi-system																				
→ xCOMET-QE																				
Vanilla	90.09	75.50	3.98	51.82	21.72	89.84	80.12	2.23	34.07	35.01	89.13	81.49	2.37	59.82	32.92	92.38	81.67	1.77	50.28	28.91
No Base	90.29	75.71	3.96	52.30	21.97	89.80	79.92	2.27	34.14	35.09	89.41	81.56	2.37	60.26	33.19	92.32	81.65	1.79	50.52	29.00
No Ref	90.25	75.55	3.94	51.98	21.80	89.99	79.98	2.24	33.79	34.86	89.32	81.58	2.35	60.08	33.13	92.22	81.33	1.81	50.05	28.77
No GPT-4	89.68	74.71	4.11	51.11	21.42	89.57	80.24	2.23	33.89	34.86	88.44	81.15	2.42	58.86	32.17	92.33	81.74	1.77	50.06	28.76
Chosen = Base	89.41	74.70	4.14	51.38	21.65	89.30	79.79	2.26	33.20	34.16	88.17	81.08	2.43	58.91	32.21	91.94	81.21	1.83	49.35	28.26
Chosen = Ref	90.21	74.68	4.05	49.70	20.34	89.52	80.41	2.23	34.42	35.20	88.04	81.06	2.41	57.73	31.20	92.35	81.94	1.75	50.12	28.73
Chosen = GPT-4	90.60	76.38	3.87	52.71	22.06	89.74	79.50	2.30	33.86	34.78	89.81	81.67	2.34	60.53	33.02	91.96	80.83	1.85	50.73	28.89
→ CometKiwi																				
Vanilla	90.25	75.82	3.94	52.21	21.98	90.09	80.45	2.22	34.23	35.16	89.26	81.70	2.34	60.01	33.06	92.44	81.93	1.77	50.49	29.03
→ chrF																				
Vanilla	90.06	74.28	4.10	48.66	19.82	89.26	80.01	2.26	34.75	35.52	87.61	80.82	2.44	56.97	30.56	92.20	81.70	1.77	50.30	28.91
No Ref	90.06	75.61	4.01	52.39	22.13	89.49	79.38	2.33	33.89	34.94	89.21	81.49	2.38	60.17	33.23	91.99	80.96	1.86	50.57	29.06
Mono-system																				
→ xCOMET-QE																				
Vanilla	89.74	74.84	4.04	51.39	21.49	89.57	80.24	2.24	33.45	34.26	88.51	81.29	2.39	59.05	32.15	92.17	81.54	1.80	49.41	28.37
CPO																				
Multi-system																				
→ xCOMET-QE																				
Vanilla	91.03	76.32	3.69	51.65	21.09	89.99	81.38	2.13	31.67	31.03	89.95	81.89	2.27	59.83	32.41	92.75	83.60	1.64	47.69	25.63
No Ref	90.50	76.60	3.75	52.96	21.86	90.32	80.72	2.24	34.00	34.28	89.91	81.86	2.31	60.59	32.77	92.44	81.97	1.78	50.67	28.55
No Base	90.80	75.98	3.79	51.96	21.67	90.09	81.40	2.09	33.20	33.49	89.59	81.73	2.30	59.94	32.82	92.74	83.13	1.65	48.54	27.17
No GPT-4	91.27	75.23	3.79	49.49	19.91	88.41	80.44	2.26	30.92	29.77	88.81	81.35	2.32	57.91	30.94	92.22	83.16	1.67	46.82	24.53
Chosen = Base	89.59	74.75	4.11	51.34	21.65	89.06	79.83	2.27	32.65	33.47	87.94	81.02	2.44	58.62	31.94	91.75	81.06	1.85	48.56	27.60
Chosen = Ref	90.71	70.87	4.05	41.49	14.40	84.25	78.61	2.40	26.14	24.09	81.95	77.86	2.53	48.75	22.02	86.97	80.01	1.79	39.81	17.62
Chosen = GPT-4	89.77	76.05	3.92	53.10	21.12	87.44	77.07	2.58	32.40	31.93	89.69	80.99	2.40	60.42	31.72	90.50	78.81	2.02	50.22	27.00
→ CometKiwi																				
Vanilla	91.03	76.69	3.61	52.68	21.40	89.35	81.27	2.23	31.90	30.95	89.82	82.04	2.24	60.22	32.58	92.19	83.64	1.71	48.11	25.35
→ chrF																				
Vanilla	89.56	68.89	4.50	38.41	12.48	85.29	79.47	2.43	28.37	26.33	78.51	75.62	2.83	45.32	19.41	88.89	80.99	1.84	42.50	19.33
No Ref	89.43	75.91	3.97	53.42	22.37	88.20	77.93	2.52	33.71	33.97	89.26	81.52	2.38	60.63	33.08	90.83	79.37	2.00	51.11	28.32
Mono-system																				
→ xCOMET-QE																				
Vanilla	90.82	76.01	3.77	51.92	21.44	90.14	81.75	2.15	33.22	33.80	89.35	81.80	2.29	59.52	32.26	92.69	82.91	1.69	49.02	27.74
Optimized	91.06	76.28	3.68	51.90	21.38	90.44	82.35	2.07	33.32	33.70	89.58	81.97	2.25	59.65	32.16	92.87	83.47	1.66	49.11	27.57

Table 9: Comprehensive downstream evaluation for the WMT'22 dataset, reporting xCOMET-QE, CometKiwi, Metric-X, chrF, and BLEU scores for all models and language pairs. Learning objectives are indicated in **bold** font, candidate settings in *italics*, and alignment metrics are preceded by an arrow (→).

	en-cs					de-en					en-de				
	Neural			Lexical		Neural			Lexical		Neural			Lexical	
	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU
Base	85.90	73.23	1.91	52.57	27.45	84.79	76.57	3.73	66.64	39.38	84.97	71.97	3.04	61.69	32.78
SFT															
Multi-system															
→ xCOMET-QE															
Vanilla	87.19	74.32	1.78	54.31	28.71	85.29	77.01	3.59	67.82	40.50	85.75	72.56	2.91	61.75	32.81
No Base	86.89	73.60	1.82	54.46	28.83	85.49	77.19	3.58	68.17	40.88	85.41	72.40	2.88	61.81	32.57
No Ref	87.44	74.13	1.83	54.43	29.20	85.46	77.19	3.64	67.90	40.46	85.78	72.71	2.91	62.05	33.15
No GPT4	87.42	74.57	1.81	53.84	28.62	84.76	76.80	3.74	67.05	39.47	85.67	72.60	2.90	61.59	32.63
Chosen = Base	86.89	74.05	1.93	52.71	27.72	84.86	76.88	3.71	67.28	39.59	85.24	72.09	2.99	61.81	32.98
Chosen = Ref	87.68	74.94	1.81	53.56	28.18	85.08	76.78	3.62	66.26	38.70	86.14	72.93	2.86	61.36	32.16
Chosen = GPT4	86.89	73.06	1.89	55.36	29.20	85.94	77.70	3.52	68.55	41.07	85.25	72.07	2.92	62.49	33.35
→ CometKiwi															
Vanilla	87.19	74.65	1.77	54.27	28.71	85.42	77.16	3.58	68.00	40.54	85.90	72.93	2.83	62.09	32.99
→ chrF															
Vanilla	87.15	74.18	1.85	53.91	28.57	84.93	76.61	3.69	65.86	38.47	85.68	72.47	2.90	61.50	32.37
No Ref	87.22	73.40	1.87	54.96	29.50	85.19	77.02	3.66	67.93	40.42	85.49	72.56	2.98	62.46	33.64
Mono-system															
→ xCOMET-QE															
Vanilla	86.89	74.78	1.86	52.57	27.60	85.37	77.03	3.64	67.64	39.86	85.54	72.80	3.03	61.82	33.10
CPO															
Multi-system															
→ xCOMET-QE															
Vanilla	86.53	76.83	1.65	49.84	23.64	86.18	77.55	3.45	67.15	39.97	87.61	73.39	2.57	58.32	27.66
No Ref	87.48	74.40	1.77	55.51	28.99	86.00	77.80	3.48	68.30	40.60	85.82	73.08	2.79	62.73	32.81
No Base	86.45	76.06	1.71	50.44	25.26	85.80	77.06	3.54	67.32	40.00	87.23	73.00	2.71	59.44	29.86
No GPT4	86.27	76.71	1.64	49.10	23.01	85.96	77.19	3.57	65.40	37.11	87.39	72.99	2.53	57.36	26.45
Chosen = Base	86.71	74.16	1.92	51.74	26.79	84.95	76.91	3.74	66.64	38.79	85.67	72.46	2.98	61.20	32.37
Chosen = Ref	72.09	72.38	1.94	37.30	12.61	82.25	74.74	4.09	55.55	24.83	86.62	69.62	2.92	49.04	17.37
Chosen = GPT4	85.67	70.85	2.06	55.59	27.58	85.41	77.18	3.53	68.10	39.75	83.74	71.58	2.92	63.11	32.22
→ CometKiwi															
Vanilla	85.70	76.97	1.73	50.28	23.74	85.76	77.66	3.47	67.43	40.07	87.52	74.21	2.57	59.08	27.90
→ chrF															
Vanilla	77.49	73.98	2.02	41.47	15.17	79.30	72.47	4.52	51.67	21.14	86.43	71.61	2.65	52.57	20.35
No Ref	85.71	71.61	2.09	56.21	28.93	85.13	77.15	3.57	68.15	40.49	84.20	71.56	3.12	63.31	33.63
Mono-system															
→ xCOMET-QE															
Vanilla	87.92	76.62	1.76	52.03	26.54	85.72	77.14	3.53	67.22	39.30	87.09	74.07	2.70	61.01	31.70
Optimized	88.39	77.36	1.71	52.11	26.38	86.00	77.76	3.55	67.05	38.79	87.32	74.64	2.57	60.71	31.07

	ru-en					en-ru					zh-en				
	Neural			Lexical		Neural			Lexical		Neural			Lexical	
	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU
Base	86.22	80.04	2.56	55.59	28.45	89.08	74.89	2.63	49.90	23.99	90.45	76.06	3.68	45.44	18.79
SFT															
Multi-system															
→ xCOMET-QE															
Vanilla	87.66	80.62	2.53	56.62	29.52	89.48	75.83	2.50	50.74	24.80	91.11	76.98	3.61	46.37	19.72
No Base	87.96	80.55	2.55	56.80	29.50	89.53	75.86	2.52	50.72	24.66	91.04	77.15	3.57	46.70	19.93
No Ref	87.79	80.64	2.54	56.64	29.34	89.23	75.43	2.58	50.88	24.82	91.16	76.92	3.58	46.52	19.95
No GPT4	86.62	80.13	2.56	55.99	29.03	89.58	75.85	2.47	50.36	24.36	90.72	76.52	3.70	45.69	19.28
Chosen = Base	86.39	80.10	2.57	55.72	28.79	88.94	74.68	2.68	50.17	24.24	90.43	76.33	3.72	45.62	18.99
Chosen = Ref	86.03	79.94	2.56	55.01	28.26	89.83	76.40	2.45	50.35	24.38	91.23	76.47	3.63	44.89	18.73
Chosen = GPT4	88.30	80.79	2.57	57.15	29.47	88.86	75.45	2.60	51.26	24.72	91.68	77.92	3.46	47.38	20.22
→ CometKiwi															
Vanilla	87.70	80.75	2.52	56.77	29.68	89.56	75.87	2.48	50.99	24.88	91.19	77.07	3.55	46.69	19.89
→ chrF															
Vanilla	85.27	79.62	2.58	54.54	28.04	89.65	75.96	2.48	50.33	24.23	91.05	76.15	3.71	44.23	18.31
No Ref	87.63	80.64	2.56	56.67	29.43	88.88	75.00	2.61	51.14	24.83	91.02	76.99	3.61	46.74	19.95
Mono-system															
→ xCOMET-QE															
Vanilla	86.94	80.33	2.56	55.86	28.64	89.16	75.09	2.64	50.02	24.10	90.84	76.61	3.64	45.89	18.97
CPO															
Multi-system															
→ xCOMET-QE															
Vanilla	88.43	80.99	2.44	56.88	29.63	91.54	79.24	2.13	48.22	21.98	91.94	77.56	3.41	46.69	19.65
No Ref	88.50	81.12	2.50	57.26	29.55	89.33	76.05	2.47	50.98	24.10	91.50	78.04	3.45	47.58	20.12
No Base	87.98	80.69	2.46	57.06	29.77	90.90	77.72	2.26	48.94	23.02	91.89	77.09	3.45	46.42	19.67
No GPT4	87.00	80.29	2.47	55.39	28.24	91.41	79.47	2.13	47.24	21.10	91.92	77.21	3.46	45.18	18.62
Chosen = Base	86.11	79.99	2.58	55.40	28.16	88.97	75.01	2.67	49.68	23.89	90.61	76.46	3.70	45.48	18.88
Chosen = Ref	77.69	76.31	2.70	46.94	20.32	89.79	76.43	2.20	39.65	14.17	90.11	72.25	3.88	39.36	14.29
Chosen = GPT4	88.19	80.22	2.65	56.56	27.95	87.25	73.65	2.78	50.57	22.96	90.68	77.67	3.59	47.74	19.60
→ CometKiwi															
Vanilla	88.34	81.08	2.44	57.08	29.59	91.08	79.22	2.21	48.53	21.83	91.71	77.91	3.39	47.51	19.89
→ chrF															
Vanilla	74.05	74.06	3.00	43.60	17.59	90.38	77.64	2.22	42.21	15.79	89.24	70.90	4.22	36.75	12.42
No Ref	87.70	80.56	2.55	57.13	29.67	87.72	74.05	2.75	51.23	23.90	90.75	77.58	3.61	47.80	20.37
Mono-system															
→ xCOMET-QE															
Vanilla	87.65	80.86	2.46	56.53	29.09	90.28	77.45	2.38	49.85	23.92	91.79	77.42	3.44	46.34	19.18
Optimized	87.63	80.84	2.45	56.56	29.03	90.64	78.18	2.28	49.94	23.82	91.79	77.58	3.39	46.59	19.31

	en-zh					xx-en					en-xx				
	Neural			Lexical		Neural			Lexical		Neural			Lexical	
	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU	xCOMET-QE	CometKiwi	Metric-X	chrF	BLEU
Base	83.92	71.42	2.56	35.58	35.94	88.00	77.74	3.23	52.30	25.37	86.19	73.08	2.42	47.31	29.43
SFT															
Multi-system															
→ xCOMET-QE															
Vanilla	84.89	72.22	2.47	36.50	37.00	88.96	78.46	3.17	53.30	26.38	87.07	73.99	2.30	48.38	30.39
No Base	84.80	72.00	2.50	36.80	37.47	89.07	78.53	3.16	53.57	26.52	86.94	73.70	2.33	48.52	30.50
No Ref	84.78	71.50	2.53	37.05	37.73	89.05	78.47	3.17	53.39	26.41	87.04	73.60	2.36	48.65	30.79
No GPT4	84.50	72.22	2.48	36.26	36.77	88.29	78.02	3.24	52.62	25.84	87.04	74.08	2.31	48.03	30.14
Chosen = Base	84.06	71.38	2.54	35.85	36.39	88.07	77.93	3.25	52.52	25.62	86.52	73.27	2.44	47.52	29.74
Chosen = Ref	84.66	72.81	2.40	36.06	36.28	88.33	77.92	3.19	51.75	25.18	87.29	74.57	2.27	47.86	29.82
Chosen = GPT4	84.75	71.33	2.54	37.51	38.34	89.57	79.06	3.11	54.08	26.67	86.70	73.18	2.39	49.23	30.97
→ CometKiwi															
Vanilla	84.90	72.44	2.48	36.43	36.92	89.03	78.57	3.13	53.53	26.53	87.11	74.21	2.29	48.45	30.40
→ chrF															
Vanilla	84.40	72.14	2.44	36.59	36.93	87.91	77.62	3.25	51.20	24.86	86.95	73.96	2.31	48.14	30.11
No Ref	84.16	70.78	2.58	37.42	38.38	88.89	78.47	3.19	53.51	26.44	86.65	73.02	2.41	49.04	31.13
Mono-system															
→ xCOMET-QE															
Vanilla	84.52	72.02	2.53	35.82	36.16	88.55	78.17	3.20	52.74	25.60	86.75	73.87	2.40	47.43	29.60
CPO															
Multi-system															
→ xCOMET-QE															
Vanilla	86.32	75.09	2.36	31.02	29.15	89.77	78.95	3.02	53.47	26.32	88.09	76.75	2.09	44.29	25.15
No Ref	85.84	73.30	2.43	36.43	36.32	89.57	79.26	3.07	54.18	26.59	87.41	74.46	2.27	48.88	30.05
No Base	85.75	74.50	2.30	32.63	31.93	89.52	78.54	3.06	53.44	26.40	87.66	75.84	2.14	45.27	26.99
No GPT4	84.69	74.65	2.45	29.78	27.30	89.16	78.46	3.07	51.94	24.91	87.45	76.62	2.11	43.30	24.02
Chosen = Base	84.64	72.42	2.50	34.26	34.58	88.05	77.95	3.25	52.24	25.22	86.68	73.75	2.41	46.54	28.75
Chosen = Ref	79.65	72.88	2.69	23.12	19.34	84.06	74.22	3.43	44.53	18.10	81.01	73.55	2.33	34.64	15.54
Chosen = GPT4	82.88	69.66	2.81	36.00	35.56	88.99	78.64	3.20	53.95	25.59	85.14	71.40	2.58	48.68	28.99
→ CometKiwi															
Vanilla	84.89	74.61	2.52	30.71	28.63	89.58	79.16	3.02	53.97	26.43	87.25	76.71	2.19	44.48	24.99
→ chrF															
Vanilla	80.91	73.49	2.69	26.29	22.83	81.79	72.38	3.77	41.46	15.65	83.21	74.76	2.34	37.96	18.13
No Ref	83.08	69.85	2.81	37.64	37.80	88.79	78.73	3.17	54.21	26.74	85.40	71.82	2.60	49.59	30.49
Mono-system															
→ xCOMET-QE															
Vanilla	85.85	74.47	2.33	34.22	34.00	89.33	78.78	3.06	53.17	25.80	87.94	76.01	2.20	46.65	28.45
Optimized	86.78	75.68	2.25	33.58	32.93	89.36	78.92	3.03	53.28	25.77	88.50	76.87	2.12	46.48	27.99

Table 10: Comprehensive downstream evaluation for the WMT'23 dataset. Metrics, notations and formatting are the same as in Table 9.

B Additional Data Details

B.1 Building Preference Datasets in the Mono-System Setting

Following the experimental setup detailed in the main text (Section 3), we here provide further details on the method used to construct mono-system preference datasets. As a reminder, after generating the K candidate translations for each source sentence, we have, for all $1 \leq i \leq N$,

$$\mathcal{Y}_i^{mono} = \{y_i^1, \cdots, y_i^K\},$$

where $y_i^1 \preceq \cdots \preceq y_i^K$ are assumed to be sorted in increasing metric score order. For each sample, we evaluate y_i^{Base} (the greedy-decoded translation) using metric m and check its rank in the set of candidate translations. We denote it by b_i. Sorted in increasing quality order, we thereby have

$$y_i^1 \preceq \cdots \preceq y_i^{b_i-1} \preceq y_i^{Base} \preceq y_i^{b_i} \preceq \cdots \preceq y_i^K.$$

Finally, to determine the chosen and rejected hypotheses, we select two offset parameters $o^r, o^c \in \mathbb{N}$, such that the chosen and rejected options are respectively

$$\begin{cases} y_i^c = y_i^{\min(K, b_i + o^c)} \\ y_i^r = y_i^{\max(1, b_i - o^r)} \end{cases}.$$

Intuitively, o^r and o^c control the average quality of the chosen and rejected options in the resulting preference dataset and ensure that the chosen (resp. rejected) option always has a higher (resp. lower) quality than the base translation. Table 11 presents the average quality properties for mono-system preference datasets, and compares them to the multi-system setting.

		Neural		Lexical
	Hyp.	xCOMET-QE	CometKiwi	chrF
Multi-system				
Candidate systems	Base	93.09	87.13	58.33
	GPT-4	94.58	88.32	60.93
	Reference	91.84	86.72	100.00
Vanilla preference dataset	Rejected	87.86	84.15	78.48
	Chosen	97.24	89.81	75.95
Mono-system				
Multi-system replica	Rejected	87.80	83.04	55.69
	Chosen	97.29	89.20	57.18
Chosen = Low / Rejected = Low	Rejected	75.36	75.46	52.95
	Chosen	93.60	87.04	57.14
Chosen = Low / Rejected = Mid	Rejected	84.54	81.02	54.93
	Chosen	93.60	87.04	57.14
Chosen = Low / Rejected = High	Rejected	92.15	85.54	55.86
	Chosen	93.60	87.04	57.14
Chosen = Mid / Rejected = Low	Rejected	75.36	75.46	52.95
	Chosen	95.77	88.40	57.43
Chosen = Mid / Rejected = Mid	Rejected	84.54	81.02	54.93
	Chosen	95.77	88.40	57.43
Chosen = Mid / Rejected = High	Rejected	92.15	85.54	55.86
	Chosen	95.77	88.40	57.43
Chosen = High / Rejected = Low	Rejected	75.36	75.46	52.95
	Chosen	98.16	89.84	57.56
Chosen = High / Rejected = Mid	Rejected	84.54	81.02	54.93
	Chosen	98.16	89.84	57.56
Chosen = High / Rejected = High	Rejected	92.15	85.54	55.86
	Chosen	98.16	89.84	57.56

Table 11: Average quality properties for xCOMET-QE-based mono-system preference datasets, compared to the multi-system setting. Multi-system replica is the mono-system dataset that matches the average chosen/rejected qualities of the multi-system preference data. Other mono-system datasets are represented by their relative average chosen/rejected qualities.

B.2 Language Statistics

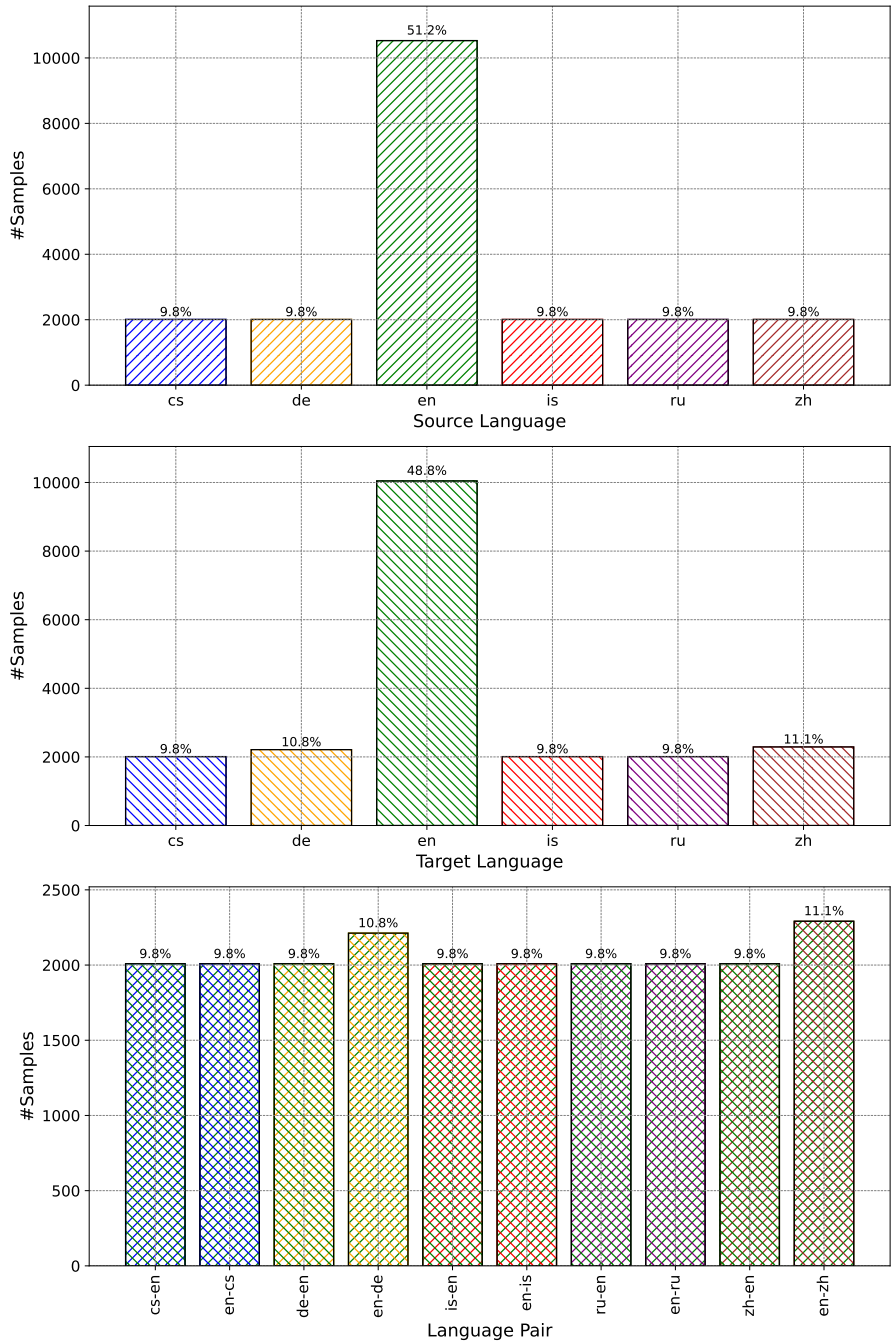

Figure 2: Language statistics for preference datasets. The y-axis represents the number of samples, corresponding percentages are displayed above each bar.

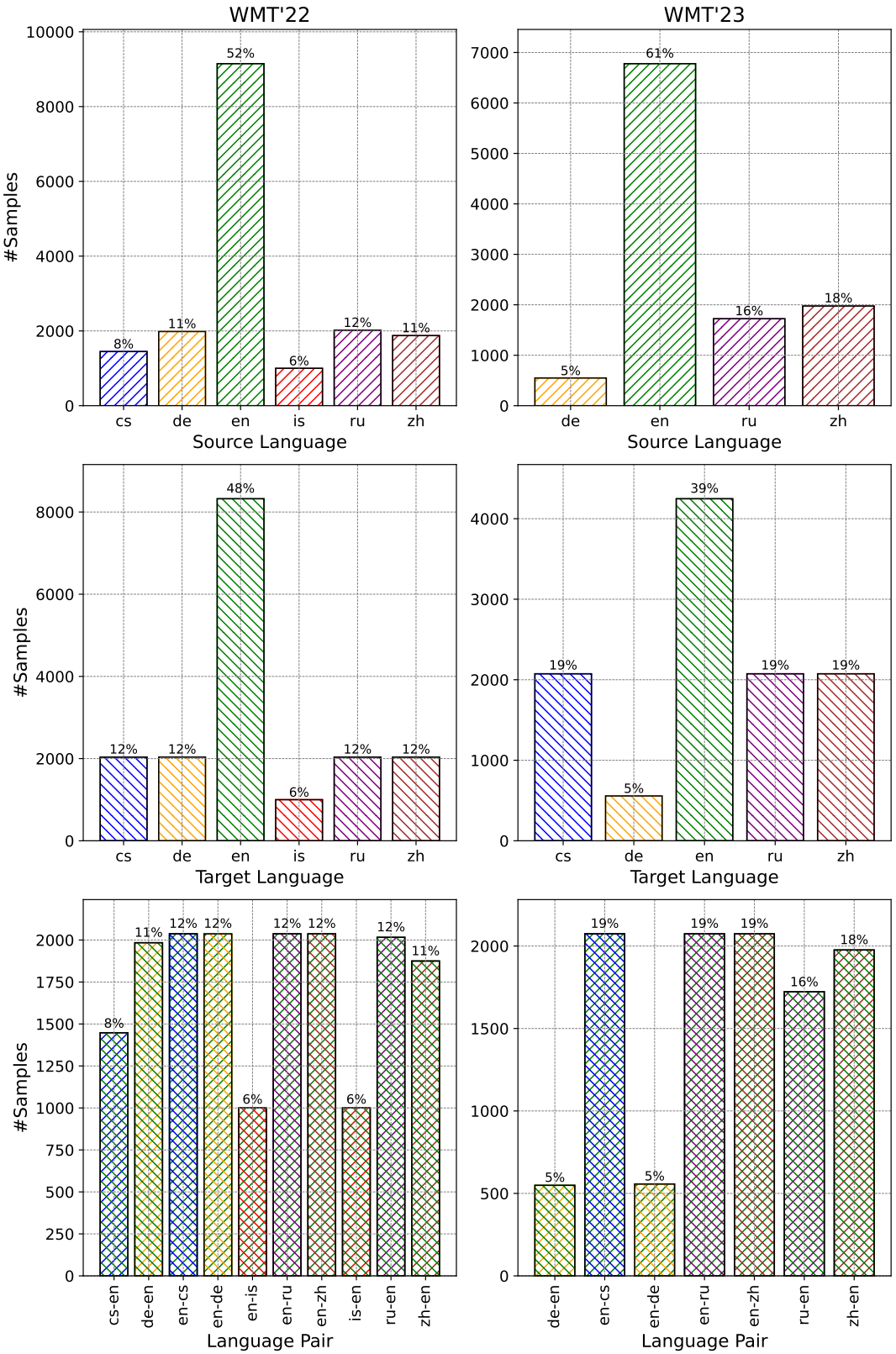

Figure 3: Language statistics for WMT'22 and WMT'23 test data. The y-axis represents the number of samples, corresponding percentages are displayed above each bar.