How Far can 100 Samples Go? Unlocking Zero-Shot Translation with Tiny Multi-Parallel Data

Di Wu   Shaomu Tan   Yan Meng   David Stap   Christof Monz
Language Technology Lab
University of Amsterdam
{d.wu, s.tan, y.meng, d.stap, c.monz}@uva.nl

Abstract

Zero-shot translation aims to translate between language pairs not seen during training in Multilingual Machine Translation (MMT) and is widely considered an open problem. A common, albeit resource-consuming, solution is to add as many related translation directions as possible to the training corpus. In this paper, we show that for an English-centric model, surprisingly large zero-shot improvements can be achieved by simply fine-tuning with a very small amount of multi-parallel data. For example, on the EC30 dataset, we obtain up to +21.7 ChrF++ non-English overall improvements (870 directions) by using only 100 multi-parallel samples while preserving English-centric translation quality. This performance exceeds M2M100 by an average of 5.9 ChrF++ in the involved non-English directions. When investigating the size effect of fine-tuning data on translation quality, we found that already a small, randomly sampled set of fine-tuning directions is sufficient to achieve comparable improvements. The resulting non-English performance is close to the complete translation upper bound. Even in a minimal setting—fine-tuning with only one single sample—the well-known off-target issue is almost completely resolved, explaining parts—but not all—of the observed improvements in translation quality.[1]

1 Introduction

The zero-shot capability shown by Multilingual Machine Translation (MMT) (Johnson et al., 2017) is of considerable significance, particularly in the context of translating between low-resource or distant language pairs. However, even for systems trained on large-scale data, the zero-shot performance is still far from sufficient (Tan and Monz, 2023), especially when scaling up the number of involved languages. Substantial efforts (Zhang et al., 2020; Pan et al., 2021; Gu and Feng, 2022; Mao et al., 2023) have been dedicated to improving the zero-shot capabilities of models trained on readily available, predominantly English-centric corpora.

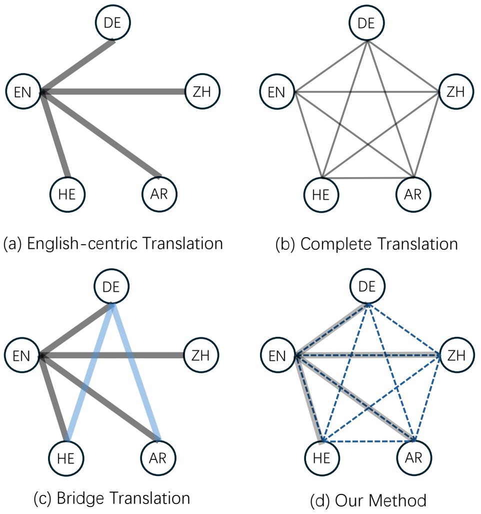


Figure 1: (a) English-centric training data is normally readily available but can only cover a few real-world directions, while (b) complete translation (Freitag and Firat, 2020) aims to cover all but suffers from the small data scale. (c) Mining partial non-English data as the bridge languages shows promising zero-shot improvements but is also resource-consuming when scaling up. (d) We show that substantial overall improvements can be achieved by fine-tuning an English-centric model with tiny extra multi-parallel data, which is readily available, like NTREX (Federmann et al., 2022).

To fully cover translation directions, Freitag and Firat (2020) propose mining multi-parallel (multi-way aligned) samples to extend the training set from English-centric to a complete multilingual one as shown in Figure 1-(b). Non-English translation quality in this setting indeed increases substantially. However, such a setting is far away from real-world practice when scaling up. As shown in Freitag and Firat (2020), to solely extend the training set from bilingual aligned to all languages (6-way) involved in their case, the amount of available data drops from 123M to 10K, which is insufficient.

[1] Models and codes are released: https://github.com/moore3930/MultiParallelFinetuning4MMT

Findings of the Association for Computational Linguistics: ACL 2024, pages 15092–15108
August 11-16, 2024 ©2024 Association for Computational Linguistics

To reflect worldwide translation needs, Fan et al. (2021) build and open-source a training dataset covering 100 languages through industry-scale mining. In addition to English-centric data, supervised data for thousands of bridge language pairs is mined and included, organized based on language families. The MMT model trained on the resulting data, M2M100, exhibits clear improvements in many non-English directions. This work drives a simple but resource-consuming solution for real-world demand: mining as much training data as possible to bridge non-English language pairs, at least at the language family level.

In this paper, we take a step back and again look at the readily available English-centric model. We empirically show that the zero-shot ability of an English-centric model can be easily unlocked via fine-tuning with a tiny amount of multi-parallel data, which is much simpler and more efficient than the extensive training data mining and bridging done by earlier work. E.g., after fine-tuning the English-centric system trained on the EC30 dataset with 100 multi-parallel samples from NTREX (Federmann et al., 2022), we observed 21.7 ChrF++ zero-shot improvements across 870 language directions. This performance surpasses M2M100 (Fan et al., 2021) by 5.9 ChrF++ (See Table 16), despite M2M100 being trained on billions of sentence pairs spanning thousands of non-English directions.

Furthermore, we investigate the size effect of fine-tuning data: 1) Surprisingly, even when fine-tuning using a randomly sampled 10% of directions, the overall improvements are nearly the same as that of full-direction fine-tuning. 2) The improvements brought by very small fine-tuning datasets only slightly lag behind the upper bound (complete translation) while preserving English-centric capabilities, showing great practical potential. 3) Even with just one single multi-parallel sample for fine-tuning, the well-known off-target problem (Zhang et al., 2020; Yang et al., 2021; Sennrich et al., 2023), is easily addressed, reducing the off-target rate from 51.8% to 1.9%. However, not all improvements in translation quality can solely be attributed to lower off-target rates as we also see clear improvements in cases where translations are already in the correct target language.

Given the high efficiency and practicality, we encourage the community to consider fine-tuning with tiny readily available multi-parallel data, like NTREX (Federmann et al., 2022), as a strong baseline for zero-shot translation. Our findings challenge some prevailing assumptions about MMT systems based on the clear evidence proposed in this paper that: 1) the off-target issue is likely overestimated and can be easily handled, and 2) the potential of single-language-centric (like English) MMT models is substantial and often overlooked. We hope these insights will inspire new discussions and explorations within the community.

2 Related Work

The zero-shot translation capability of MMT is associated with multilingualism, following the hypothesis of universal representation or interlingua. Arivazhagan et al. (2019) view zero-shot translation as a domain adaptation problem (Ben-David et al., 2006) in MMT, and apply auxiliary losses to explicitly incentivize the model to learn and use domain- (language-) invariant representation. Liu et al. (2021) attribute the low quality of zero-shot MT to the positional correspondence to input tokens, which hinders modeling language-agnostic representation. Pan et al. (2021) and Stap et al. (2023) use a contrastive loss to close the representational gap between different languages. Some other approaches aim to harness the capabilities of pretrained multilingual models for zero-shot translation. Chen et al. (2022) employ multilingual pretrained encoders to extend bilingual translation to many-to-one translation. Recently, some work has focused on leveraging pre-trained large language models for multilingual translation (Zhang et al., 2023; Moslem et al., 2023). Despite the inclusion of the so-called "emergent abilities" (Wei et al., 2022) triggered by zero-shot prompting, we categorize these works as following a similar line.

This paper focuses on a data-centric approach for comprehensively improving zero-shot performance. We empirically show that a well-trained English-centric model can be easily boosted for overall zero-shot capability via fine-tuning with minimal data, even if only covering a small portion of translation directions (10%). This allows us to leverage multi-parallel data, which is hard to obtain in large quantities, resulting in a highly efficient and practical solution to overall zero-shot translation. We note that Maillard et al. (2023) also show that integrating small high-quality data (6K samples) into the training corpus can have a big impact on low-resource translation systems, especially when combined with back translation (Sennrich et al., 2016). However, we contend that the reasons for the effectiveness differ: 1) as shown in Section 3.6, the substantial improvements persist when using data built from the training set, where the influence from domain or quality level is eliminated, and 2) our method can work with extremely minimal fine-tuning data (100 or even a single sample).

3 Experiments

In this paper, we propose a simple approach that largely improves overall zero-shot translation quality, i.e., fine-tuning an English-centric model in multiple directions with sentence pairs constructed from small, readily available multi-parallel datasets, like NTREX (Federmann et al., 2022). We refer to such a process as “fine-tuning with multi-parallel data” or “multi-parallel fine-tuning”.

3.1 Fine-Tuning Data Construction

Given a multi-parallel dataset comprising N distinct languages, each with K samples, we can generate pairwise data in all $N \times (N-1)$ possible directions. Note that acquiring large quantities of multi-parallel data poses challenges due to many professional human translators being involved. However, horizontally expanding a readily available multi-parallel dataset to include one more language is straightforward. It simply requires annotating K additional samples for the new language based on the current dataset, with $2 \times N$ new translation directions indirectly covered.

3.2 Datasets

NTREX-128. NTREX[2] (Federmann et al., 2022) is initially proposed as an evaluation dataset, expanding multilingual testing for translation from English into 128 target languages, which consists of 1997 samples per language and mainly focus on the News domain. Given the multi-parallel organization of NTREX data, we can easily build arbitrary pairwise data across 128 languages. In this paper, we leverage NTREX to create our fine-tuning datasets and conduct experiments to highlight the big impact of such a tiny amount of data.

Europarl-8. Europarl[3] (Koehn, 2005) consists of 20 English-centric language pairs from the proceedings of the European Parliament, with sizes ranging from 399K to 2M. A characteristic of Europarl is that part of the samples are multi-way aligned. In this paper, we select the eight most resource-rich languages, i.e., EN, DA, DE, ES, FI, FR, IT, and NL, to mine a fully multi-parallel dataset named Europarl-8 via aligning multi-way sentences with the same English part. This results in about 1.2M fully multi-parallel data instances, where each sentence has 7 counterparts in other languages.

EC30. To ensure a more diverse and inclusive large-scale evaluation, we follow Tan and Monz (2023); Wu and Monz (2023) and use the EC30 dataset, which is built from WMT (Bojar et al., 2017) and OPUS (Tiedemann, 2012) corpora. EC30 comprises 61 million English-centric bilingual sentences for training, encompassing 30 non-English languages with diverse resource levels (High: 5M, Medium: 1M, Low: 100K). Each resource group includes languages from 5 families with multiple writing systems.

Evaluation Benchmark. For all of the experiments in this paper, we evaluate translations via the Flores-101 benchmark (Goyal et al., 2022). Flores comprises 3001 sentences sourced from English Wikipedia, which covers a variety of topics and domains and is translated into 101 languages by professional translators. We use *dev* and *devtest* as the validation and test dataset, consisting of 997 and 1012 samples, respectively. All results are evaluated on three widely used metrics, namely, ChrF++ (Popović, 2017), SacreBLEU (Post, 2018)[4], and COMET (Rei et al., 2020), to demonstrate the consistency of improvements across a broad spectrum of evaluation metrics. A more detailed description of the datasets is provided in Appendix A.1.

3.3 Experimental Setup

3.3.1 Training Setting

For experiments on the EC30 dataset, we use Transformer-Big with 16 attention heads, 1,024 embedding dimensions, and 4,096 feedforward dimensions. For Europarl-8, we use a smaller backbone, as the training data is smaller, where a standard 6-layer encoder, 6-layer decoder transformer model is applied with 4 attention heads, 512 embedding dimensions, and 1,024 feedforward dimensions. In total, 447M and 64M training parameters are involved for the two models. More detailed training settings are provided in Appendix A.2.

[2] `https://github.com/MicrosoftTranslator/NTREX`

[3] `https://www.statmt.org/europarl`

[4] nrefs:1|case:mixed|eff:no|tok:13a|smooth:exp|version:2.3.1

	Model	H-H	H-M	H-L	M-H	M-M	M-L	L-H	L-M	L-L	AVG
Two-Tag	Baseline	11.0	14.8	10.6	11.3	14.9	10.5	13.7	17.4	10.9	12.8
	Boost-100	37.9	39.1	30.6	38.0	38.8	30.5	33.9	34.8	26.8	34.5
	Boost-All	38.6	39.9	32.0	38.7	39.5	31.8	34.9	35.9	28.4	35.5
	Δ-100	+26.9	+24.3	+20.0	+26.7	+23.9	+20.0	+20.2	+17.4	+15.9	+21.7
	Δ-All	+27.6	+25.1	+21.4	+27.4	+24.6	+21.3	+21.2	+18.5	+17.5	+22.7
One-Tag	Baseline	28.0	30.4	20.0	27.9	29.5	19.7	24.0	26.0	16.4	24.7
	Boost-100	36.6	37.4	29.0	36.5	36.9	28.8	32.1	32.7	24.8	32.8
	Boost-All	37.2	38.2	30.6	37.3	37.8	30.5	33.4	34.0	26.9	34.0
	Δ-100	+8.6	+7.0	+9.0	+8.6	+7.4	+9.1	+8.1	+6.7	+8.4	+8.1
	Δ-All	+9.2	+7.8	+10.6	+9.4	+8.3	+10.8	+9.4	+8.0	+10.5	+9.3

Table 1: Zero-shot performance (ChrF++, 870 directions) trained on the EC30 dataset (61M English-centric sentence pairs), grouped by **H**igh-, **M**edium, and **L**ow-resource, respectively. Δ-100 and Δ-All mean the corresponding performance changes after fine-tuning compared to the baselines. Large improvements are observed in both One-Tag and Two-Tag settings after fine-tuning with 100 samples. Results in SacreBLEU and COMET are provided in Table 12 and Table 14, respectively. The comparison with M2M100 is provided in Table 16.

	Model	High		Medium		Low		Overall		
		EN-X	X-EN	EN-X	X-EN	EN-X	X-EN	EN-X	X-EN	AVG
Two-Tag	Baseline	52.5	57.5	53.9	56.5	42.5	49.8	49.6	54.6	52.1
	Boost-100	51.9	56.6	53.6	57.0	42.6	50.2	49.4	54.6	52.0
	Boost-All	51.9	56.0	53.6	56.6	43.2	50.9	49.6	54.5	52.0
	Δ-100	-0.6	-0.9	-0.3	+0.5	+0.1	+0.4	-0.2	0.0	-0.1
	Δ-All	-0.6	-1.5	-0.3	+0.1	+0.7	+1.1	0.0	-0.1	-0.1
One-Tag	Baseline	52.6	57.0	54.0	56.2	42.9	49.6	49.8	54.3	52.1
	Boost-100	52.0	56.1	53.7	56.5	43.2	49.9	49.6	54.2	51.9
	Boost-All	51.7	55.7	53.5	56.2	43.5	50.3	49.6	54.1	51.8
	Δ-100	-0.6	-0.9	-0.3	+0.3	+0.3	+0.3	-0.2	-0.1	-0.2
	Δ-All	-0.9	-1.3	-0.5	0.0	+0.6	+0.7	-0.3	-0.2	-0.3

Table 2: English-centric performance (ChrF++, 60 directions) trained on the EC30 dataset (61M English-centric sentence pairs). EN-X and X-EN denote the average out-of- and into-English translation performance for each resource group, respectively. It's easy to see that the performance changes (Δ-100 and Δ-All) after fine-tuning are negligible. Results in SacreBLEU and COMET are provided in Table 13 and Table 15, respectively.

3.3.2 Fine-Tuning Setting

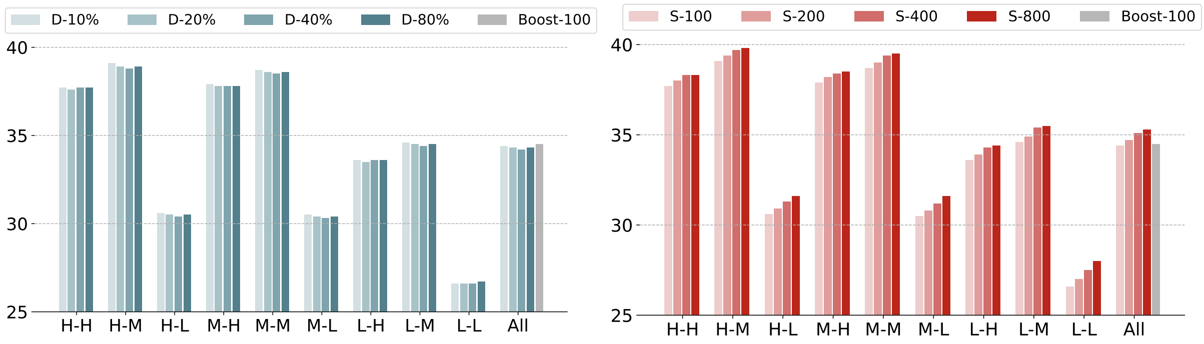
We use full-parameter fine-tuning and keep our setup as simple as possible to highlight generalizability. We reset all running statuses, including optimizer, lr scheduler, and data loaders. Also, the fine-tuning parameters are aligned with those in the training period, except for the experiments in Section 4, where we set batch accumulation to 1 as extremely small fine-tuning data is used.

Note that, because we only fine-tune with a small amount of data, the process is highly efficient: most fine-tuning experiments in this paper were completed within 1 GPU hour.

3.4 Large-Scale Experiments on EC30

In this section, we show how far a tiny amount of multi-parallel data can improve the zero-shot capability of an already well-trained large-scale English-centric MMT system. We conduct experiments on EC30, involving 30 English-centric and 870 zero-shot directions. We build fine-tuning data based on NTREX to cover all except Occitan-related directions—Occitan is not included in NTREX—as described in Section 3.2. It is noteworthy that MMT systems typically use two language tag strategies: 1) the one-tag strategy, i.e., adding the target language IDs to the encoder input, which is shown by Wu et al. (2021) to be more effective for zero-shot translation, or 2) the two-tag strategy, i.e., adding source and target language IDs to the encoder and decoder input, respectively, which is often applied to recent large-scale MMT systems (Fan et al., 2021; Pan et al., 2021; Costa-jussà et al., 2022). To show comprehensive results, we conducted experiments in both settings.

Table 1 shows the zero-shot performance across 9 resource groups. Boost-All refers to using all 1997 multi-parallel samples from NTREX to con-



(a) Increasing the number of directions (D)

(b) Increasing the number of samples (S)

Figure 2: Zero-shot performance (ChrF++) on EC30 for each scaling step, grouped by **H**igh-, **M**edium, and **L**ow-resource, respectively. (a) When we randomly selected {10%, 20%, 40%, 80%} of directions and fine-tuned with 100 samples, the overall zero-shot performance stayed nearly unchanged. However, (b) when we fixed 10% of the directions and increased the fine-tuning samples from 100 to 800, we observed consistent improvements across all resource groups. Note that S-100 and D-10% are identical. Meanwhile, the directions or samples involved in the previous step are always subsets of those in the subsequent step.

struct pair-wise fine-tuning data for all directions (including English-centric ones), while Boost-100 means using only 100 randomly sampled samples, rather than 1997, to construct the fine-tuning data.

We find that 1) fine-tuning with tiny data leads to very strong overall improvements for both tagging strategies, with up to +9.3 and +22.7 average ChrF++ point gains, respectively. 2) The zero-shot capability of the two-tag baseline lags behind the one-tag baseline, in line with Wu et al. (2021). However, after fine-tuning with multi-parallel data, the overall performance in the two-tag setting consistently outperforms the one-tag setting for each group, yielding an average margin of +1.5 ChrF++ (35.5 v.s., 34.0). Consistent improvements also hold for other metrics, see Appendix A.3.

In Table 6, we show that 854 out of 870 zero-shot directions get strong boosts (more than 10.0 ChrF++). The resulting zero-shot performance even averagely surpasses the industrial system, M2M100, by 5.9 ChrF++ (See Table 16), despite M2M100 being trained on billions of sentence pairs spanning thousands of non-English directions.

In Table 2, we show the impact of fine-tuning on English-centric directions: The trade-off effect mainly occurs in the high-resource group. However, the influence on the medium- and low-resource groups is negligible or even positive, especially for the low-resource part, resulting in nearly unchanged overall English-centric performance. For instance, fine-tuning with 100 multi-parallel samples on the two-tag model yields +21.7 ChrF++ zero-shot gains, with negligible drops in averaged English-centric performance (-0.1 ChrF++).

It is noteworthy that fine-tuning with just 100 samples achieves comparable improvements to using the entire NTREX dataset (+21.7 v.s. +22.7 in Table 1), even though the latter's size is 20 times larger. This diminishing effectiveness naturally leads us to ask (i) whether more fine-tuning data or more fine-tuning directions is important and (ii) how close can our method come to the upper-bound improvements.

We answer both questions in Section 3.5 and 3.6, respectively. If not specified, we employ the two-tag strategy in subsequent experiments because of its higher zero-shot and English-centric performance after fine-tuning.

3.5 More Data or More Directions?

In Section 3.4, we showed that fine-tuning an English-centric model with a small amount of bitext derived from NTREX (covering all directions) yields substantial zero-shot improvements. A natural assumption is that the improvement in each direction is triggered by the corresponding directional data. In this section, we investigate whether this is true, i.e., what will happen if we only cover a subset of translation directions during fine-tuning.

We conduct experiments on the same English-centric model trained on EC30, see Section 3.4, and control the scale of the fine-tuning data in the following two settings: (a) We conducted a random sampling of 100 multi-parallel NTREX sentences to construct pairwise data to cover all 870 direc-

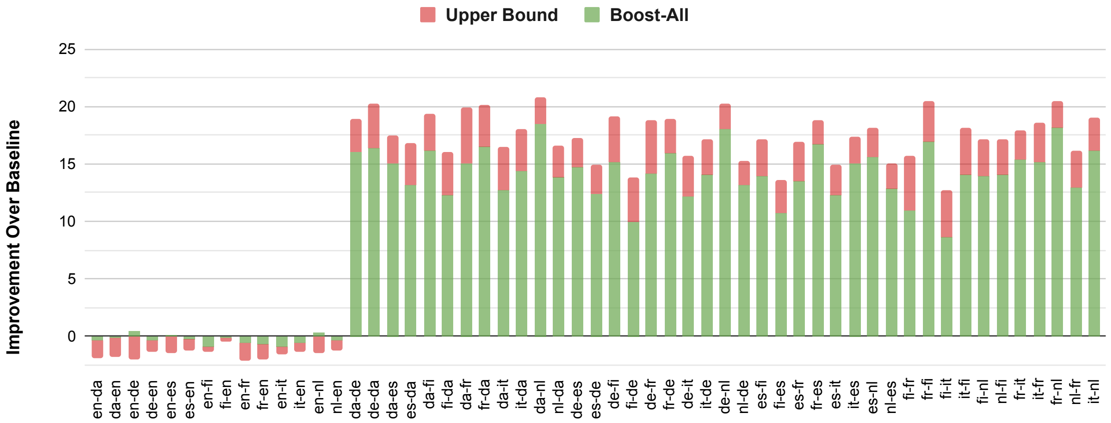


Figure 3: ChrF++ improvements of the upper bound and boosted models over the English-centric baseline on the Europarl-8 dataset. It is clear that the overall non-English capability of the boosted model is close to the upper bound (complete translation), meanwhile, it also holds the performance in English-centric directions.

tions[5]. Then, we randomly sampled {10%, 20%, 40%, 80%} directions for fine-tuning. (b) We fixed the 10% of directions as mentioned in (a) and conducted a random sampling of {100, 200, 400, 800} multi-parallel NTREX instances to construct the fine-tuning set for the corresponding directions.

Note that the bitext size in settings (a) and (b) for each scaling step is kept identical, e.g., to facilitate a fair comparison with the setting of fine-tuning with 100 multi-parallel samples in 80% directions, we also consider fine-tuning in 10% directions with 800 multi-parallel samples. Meanwhile, the directions or samples involved in the previous step are always subsets of those in the subsequent step.

In Figure 2, we show all of the corresponding fine-tuning results. Surprisingly, when fixing the size of the multi-parallel samples to 100 and then increasing the fine-tuning directions from 10% to 80%, no improvement is observed for any resource group (Figure 2-a). Fine-tuning in randomly sampled 10% directions using 100 samples achieves comparable overall results to fine-tuning in all directions (Boost-100). However, when we fix the directions to 10% and increase the multi-parallel sample size from 100 to 800, consistent improvements for all groups can be observed (Figure 2-b). This shows that the overall improvements are not sensitive to the number of directions, at least when the directions extend to a certain scale, like 10%.

In Appendix A.4.1, we further show that when we limited the fine-tuning direction set to fall in a specific family (Germanic), overall improvements only slightly lag behind that of fully fine-tuning, showing surprising boosting effects.

3.6 How Close to the Upper Bound?

In this section, we show to what extent our fine-tuning method can approximate an upper bound. Here, we consider the performance of complete translation, i.e., training with fully multi-parallel data, as the "upper bound", since identical scales of non-English bitext cover all the directions that the English-centric counterparts cannot cover.

We conduct experiments on Europarl-8, see Section 3.2, where 8-way aligned data are available. Both English-centric and complete translation models are trained based on it. Note that we reuse the former's vocabulary for the latter to ensure a fair comparison. Also, we present the results after fine-tuning the English-centric model using full-direction pairwise data constructed from all of the 1,997 samples from NTREX.

In Figure 3, we show that 1) the upper-bound performance, i.e., that of the complete model, surpasses the baseline by a large margin, resulting in +17.4 average ChrF++ gains for non-English directions. 2) However, the boosted model's performance closely approaches the upper bound in all non-English directions. 3) For the 14 English-centric directions, both the upper bound and the boosted models exhibit degradation compared to the baseline, which reveals trade-off effects from English-centric to non-English directions. However, the boosted method only slightly degrades for a few English-centric directions (e.g., en-fi and en-it in Figure 3), whereas the upper bound model

[5] These directions include English-centric ones while excluding Occitan-related ones as it is not available in NTREX.

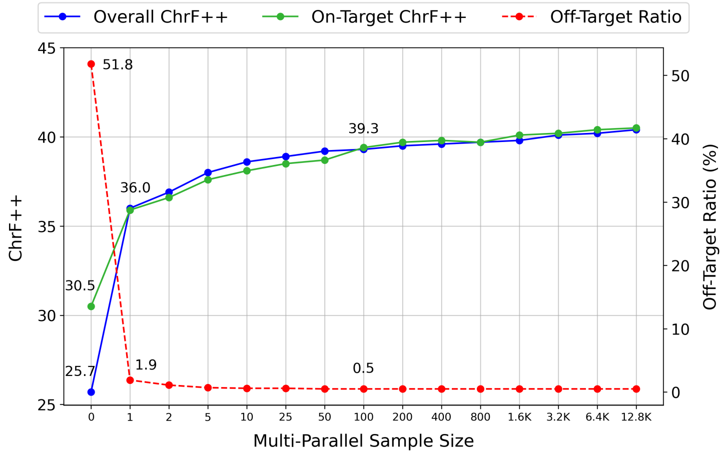


Figure 4: Zero-shot performance and off-target ratio on Europarl-8 at each scaling step. Surprisingly, fine-tuning with one sample almost completely handles the off-target problem (from 51.8 % to 1.9%). The green solid line denotes the quality improvements of the translation samples that have no off-target issue, where the gains from one single sample are also large (+5.5 ChrF++). Mean and variance for each scaling step are provided in Table 17.

drops for most. Detailed scores, including those in other metrics, are provided in Appendix A.5. In short, the boosted model achieves strong non-English gains (+14.3 ChrF++, 42 directions) with a negligible cost in English-centric directions.

4 Analysis

4.1 Off-Target and Fine-Tuning Data Size

The off-target problem has been viewed as a primary cause that impairs the zero-shot capability (Zhang et al., 2020; Yang et al., 2021; Sennrich et al., 2023; Chen et al., 2023). In this section, we delve into the impact of fine-tuning data size on off-target ratios and final performance. Moreover, we disentangle the gains of the already on-target translations, showing the extent to which the enhancements are beyond alleviating the off-target issue. Note that since Europarl-8 is fully multi-parallel, we can readily build the corresponding full-direction fine-tuning data at different scales.

Here, we sample multiple sets of multi-parallel instances from the training set of Europarl-8, ranging from 1 to 12.8K with different seeds 3 times. The average results in ChrF++ after fine-tuning for each scaling step are provided in Figure 4. Also, we report the corresponding off-target ratio evaluated by fastText[6] (Joulin et al., 2017) following previous works (Yang et al., 2021; Costa-jussà et al., 2022).

In Figure 4, the blue solid line shows the overall zero-shot performance at each scaling step, where the starting point (fine-tuning with 0 samples) denotes the performance of the original English-centric model. Notably, a high off-target ratio (51.8%) exists at this point. Surprisingly, even fine-tuning with just one multi-parallel sample, very strong overall zero-shot improvements can be obtained (from 25.7 to 36.0 ChrF++). Meanwhile, the off-target issue is almost completely resolved, dropping from 51.8% to 1.9%. Increasing from 1 to 100 samples, we can still observe clear zero-shot capability boosting (from 36.0 to 39.3 ChrF++), while the off-target change is marginal. Further scaling up fine-tuning data from the point of 100 samples shows nearly linear performance gains.

We further disentangle the evaluation set into on-target parts[7] for each language direction, where each source sample is already translated into the correct direction when evaluating the English-centric model. In Figure 4, the green solid line denotes the average performance on the on-target part. It is easy to see that the improvements brought from one single fine-tuning sample are still strong (+5.5 ChrF++), even after isolating the impact of off-target issues. As the number of fine-tuning samples increases, on-target improvements closely follow the trend of the overall improvements, further showing that the overall improvements are not only due to resolving the off-target issue.

Given the clear findings mentioned above, we question previous works (Zhang et al., 2020; Sennrich et al., 2023; Chen et al., 2023) that attribute off-target as a challenging problem and primary source for the low zero-shot performance.

[6] https://github.com/facebookresearch/fastText

[7] Note that the on-target sample size varies across directions, with an average of 488 samples per direction.

Model	H-H	H-M	H-L	M-H	M-M	M-L	L-H	L-M	L-L	ZS-AVG	EN-AVG
Baseline	11.0	14.8	10.6	11.3	14.9	10.5	13.7	17.4	10.9	12.8	52.1
(a) Multi-Parallel	37.6	38.9	30.1	36.7	37.3	29.0	32.1	32.9	24.8	33.3	50.9
(b) Multi-Directional	37.6	39.1	30.3	37.1	38.0	29.7	33.0	34.1	26.0	33.9	51.5

Table 3: Decoupling multi-parallel and multi-directional fine-tuning on the EC30 dataset. ZS-AVG and EN-AVG denote the average results of the zero-shot and English-centric performance in ChrF++, respectively. In all groups, the performance in setting (a) closely trails but never surpasses that in setting (b), showing that the boosting effects do not depend on multi-way semantic equivalence.

4.2 Does Multi-Parallelism Matter?

We have shown remarkable boosting effects obtained from a tiny amount of multi-parallel data. But, does the data have to be *multi*-parallel? In this section, we explore whether utilizing multi-parallel data, instead of just pairwise data, for fine-tuning is vital for significant enhancements.

To this end, we design experiments to decouple the impacts of multi-parallel signals. Let's consider creating fine-tuning data in 5 languages from NTREX, involving 20 directions. We carefully control the resulting bitext distribution: Firstly, we randomly map the 1,997 multi-parallel samples into 10 buckets, ensuring an even allocation of approximately 100 samples per bucket. Then, we construct pairwise data in the following two ways:

(a) Multi-Parallel: We constructed pairwise samples in the 20 directions using the multi-parallel data in one randomly picked bucket.

(b) Multi-Directional: For each bucket, we construct fine-tuning samples for only one specific language pair (2 directions), also resulting in the 20 translation directions.

Note that the size of the bitext in settings (a) and (b) are identical. In (a), each sentence has semantically equivalent counterparts in all other 4 languages. However, in (b), each sentence has only one counterpart, resulting in simple pairwise data.

To cover different language families and resource levels, we choose DE, FR, RU, HE, and AR as these 5 languages. In Table 3, we show the results that fine-tune the EC30-based English-centric model with data in (a) multi-parallel and (b) multi-directional settings, respectively. Firstly, compared to the baseline model, clear improvements can be observed for zero-shot translation in both settings. Meanwhile, in all groups, the performance in setting (a) closely trails but never surpasses that in setting (b). It shows that the boosting effects do NOT depend on multi-way semantic equivalence, implying simple multi-directional data is sufficient in case fully multi-parallel samples do not exist.

Setting	EN-X	X-EN	Zero-Shot	Off-Target (%)
Baseline	49.8	51.3	25.7	51.8
Numbers	49.9	51.5	26.8	46.1
Words	48.3	50.7	35.8	3.6
NTREX	49.5	50.9	40.0	0.5

Table 4: The results (ChrF++) after fine-tuning. "Numbers", "Words", and "NTREX" denote different types of fine-tuning data (see Section 4.3).

4.3 The Role of Semantic and Syntactic Information in Fine-Tuning Data

Considering that a small amount of fine-tuning data, e.g., 100 or even one single sample, can still substantially enhance overall zero-shot performance, a related question arises: To what extent do these improvements stem from the intrinsic information inherent in the data itself? In this section, we provide some insights into the role that the semantics and syntactic of fine-tuning data play in the surprising improvements for zero-shot translation.

We choose the English-centric model trained on Europarl-8 as our baseline (see Section 3.6), and fine-tune it on three datasets as follows:

Number Pairs. For each direction, we perform a uniform sampling of digits (ranging from 1 to 1000) multiple times, concatenating them to a certain length. Then, it is replicated for both the source and target sides, forming a number translation sample, as shown in Figure 6. Given no semantic other than numerical information is contained in this setting, we try to check whether the improvements stem from factors other than the data itself, e.g., the tags.

Word Pairs. We utilize bilingual dictionaries from MUSE [8] (Lample et al., 2018) to build word pairs for all of the directions. MUSE contains 110 English-centric bilingual dictionaries and all languages of Europarl-8 are included. We first select

[8] https://github.com/facebookresearch/MUSE

the intersection of English words for the 7 involved English-centric dictionaries. Then, we extend them by mapping words paired with the same English words together[9]. E.g., given an EN-DE pair {*bike*, *Fahrrad*} and an EN-NL pair {*bike*, *fiets*}, we can build a new DE-NL word pair {*Fahrrad*, *fiets*}. Finally, we built 28 dictionaries (16,737 word pairs for each) covering all 56 directions.

Sentence Pairs. We use 100 randomly selected multi-parallel samples from NTREX to construct pairwise data covering all directions.

To ensure a fair comparison, we maintain similar surface information across the three datasets, such as aligning the number of tokens in the number-pair and word-pair datasets with the English portion in the sentence-pair dataset. Table 4 shows the corresponding fine-tuning results: 1) Fine-tuning with number pairs results in marginal improvements. Conversely, fine-tuning with word pairs leads to noticeable zero-shot improvements (+10.1 ChrF++). Simultaneously, the off-target ratio also decreases to an acceptable level. This means that semantic information, particularly at the lexical level, plays an important role here. 2) When using sentence-pair data (NTREX) to fine-tune, considerable further improvements compared to word-pair counterparts can be observed, showing that syntactic-level information also matters.

5 Conclusion

In this paper, we show that the zero-shot performance of an English-centric MMT model can be easily boosted by a tiny amount of multi-parallel data. On EC30, +21.7 ChrF++ average gains can be achieved by fine-tuning using 100 samples from NTREX, meanwhile preserving the English-centric performance, see Section 3.4. More surprisingly, we show that fine-tuning on a small portion (10%) of directions can achieve comparable improvements to full-direction fine-tuning, see Section 3.5, which are even close to the ideal but impractical upper-bound model, see Section 3.6.

In terms of using language tags, we show that fine-tuning can address the two-tag model's performance degradation in zero-shot directions (Wu et al., 2021). Moreover, the final performance substantially surpasses that of the one-tag model across multiple metrics, see Section 3.4.

We also question earlier findings (Zhang et al., 2020; Yang et al., 2021; Sennrich et al., 2023; Chen et al., 2023) that consider the off-target issue as a challenging problem for MMT. This paper shows that the off-target issue can be easily addressed by fine-tuning with even a single sample. Moreover, we also observe clear gains on the already on-target translations after fine-tuning with a few samples, implying that the off-target issue is not the primary source, but more like a symptom, of low zero-shot quality, see Section 4.1.

By decoupling the impacts of multi-parallel signals, we demonstrate that fine-tuning data does not have to be *multi*-parallel to achieve significant improvements, see Section 4.2. However, we still recommend using multi-parallel data, as it is already sufficient, albeit small-scale, and convenient for building fine-tuning data in each involved direction. Lastly, we shed some light on the impact of different types of signals from the fine-tuning data on the final performance, see Section 4.3.

Given the clear advantages of our proposed method, we encourage the community to consider fine-tuning as a strong baseline for zero-shot translation, especially in the two-tag setting. Furthermore, we suggest that the community reconsider the real bottlenecks of MMT systems in light of the evidence presented in this paper: 1) the off-target issue is overestimated and can be easily managed, and 2) the potential of single-language-centric (like English) MMT models, is substantial and often overlooked.

Limitations

Multi-parallel data are normally built in a way that translates the same English data into multiple other languages by professional human translators. Hence, in the resulting non-English fine-tuning data, both the source and target side are translated instead of using the original text. This may exacerbate potential drawbacks in certain directions, such as translationese.

Broader Impact

MMT systems have significant progress recently. However, potential challenges such as mistranslation or fairness problems still exist, e.g., the generation ability is not guaranteed to be fair across languages or demographic features, which may run the risk of reinforcing societal biases.

[9] Specifically, for each one-to-many mapping that exists, we randomly select a one-to-one mapping.

Acknowledgement

We thank the anonymous reviewers for their efforts to make this paper better. We thank our colleagues from the Language Technology Lab at the University of Amsterdam, especially Seth Aycock, Sergey Troshin, and Wafaa Mohammed for their suggestions on writing. We thank Chongyang for its invaluable spiritual support. This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under project number VI.C.192.080.

References

Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Roee Aharoni, Melvin Johnson, and Wolfgang Macherey. 2019. The missing ingredient in zero-shot neural machine translation. *arXiv preprint arXiv:1903.07091*.

Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. 2006. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Shujian Huang, Matthias Huck, Philipp Koehn, Qun Liu, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Raphael Rubino, Lucia Specia, and Marco Turchi. 2017. Findings of the 2017 conference on machine translation (WMT17). In *Proceedings of the Second Conference on Machine Translation*, pages 169–214, Copenhagen, Denmark. Association for Computational Linguistics.

Guanhua Chen, Shuming Ma, Yun Chen, Dongdong Zhang, Jia Pan, Wenping Wang, and Furu Wei. 2022. Towards making the most of cross-lingual transfer for zero-shot neural machine translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 142–157, Dublin, Ireland. Association for Computational Linguistics.

Liang Chen, Shuming Ma, Dongdong Zhang, Furu Wei, and Baobao Chang. 2023. On the off-target problem of zero-shot multilingual neural machine translation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9542–9558, Toronto, Canada. Association for Computational Linguistics.

Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, et al. 2021. Beyond english-centric multilingual machine translation. *The Journal of Machine Learning Research*, 22(1):4839–4886.

Christian Federmann, Tom Kocmi, and Ying Xin. 2022. NTREX-128 – news test references for MT evaluation of 128 languages. In *Proceedings of the First Workshop on Scaling Up Multilingual Evaluation*, pages 21–24, Online. Association for Computational Linguistics.

Markus Freitag and Orhan Firat. 2020. Complete multilingual neural machine translation. In *Proceedings of the Fifth Conference on Machine Translation*, pages 550–560, Online. Association for Computational Linguistics.

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.

Shuhao Gu and Yang Feng. 2022. Improving zero-shot multilingual translation with universal representations and cross-mapping. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6492–6504, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain. Association for Computational Linguistics.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand.

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

Danni Liu, Jan Niehues, James Cross, Francisco Guzmán, and Xian Li. 2021. Improving zero-shot translation by disentangling positional information. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages

1259–1273, Online. Association for Computational Linguistics.

Jean Maillard, Cynthia Gao, Elahe Kalbassi, Kaushik Ram Sadagopan, Vedanuj Goswami, Philipp Koehn, Angela Fan, and Francisco Guzman. 2023. Small data, big impact: Leveraging minimal data for effective machine translation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2740–2756, Toronto, Canada. Association for Computational Linguistics.

Zhuoyuan Mao, Raj Dabre, Qianying Liu, Haiyue Song, Chenhui Chu, and Sadao Kurohashi. 2023. Exploring the impact of layer normalization for zero-shot neural machine translation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1300–1316, Toronto, Canada. Association for Computational Linguistics.

Yasmin Moslem, Rejwanul Haque, and Andy Way. 2023. Adaptive machine translation with large language models. *arXiv preprint arXiv:2301.13294*.

Xiao Pan, Mingxuan Wang, Liwei Wu, and Lei Li. 2021. Contrastive learning for many-to-many multilingual neural machine translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 244–258, Online. 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.

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

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.

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.

Rico Sennrich, Jannis Vamvas, and Alireza Mohammadshahi. 2023. Mitigating hallucinations and off-target machine translation with source-contrastive and language-contrastive decoding. *arXiv preprint arXiv:2309.07098*.

David Stap, Vlad Niculae, and Christof Monz. 2023. Viewing knowledge transfer in multilingual machine translation through a representational lens. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 14973–14987, Singapore. Association for Computational Linguistics.

Shaomu Tan and Christof Monz. 2023. Towards a better understanding of variations in zero-shot neural machine translation performance. *arXiv preprint arXiv:2310.10385*.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218, Istanbul, Turkey. European Language Resources Association (ELRA).

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. 2022. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.

Di Wu and Christof Monz. 2023. Beyond shared vocabulary: Increasing representational word similarities across languages for multilingual machine translation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9749–9764, Singapore. Association for Computational Linguistics.

Liwei Wu, Shanbo Cheng, Mingxuan Wang, and Lei Li. 2021. Language tags matter for zero-shot neural machine translation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3001–3007, Online. Association for Computational Linguistics.

Yilin Yang, Akiko Eriguchi, Alexandre Muzio, Prasad Tadepalli, Stefan Lee, and Hany Hassan. 2021. Improving multilingual translation by representation and gradient regularization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7266–7279, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

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.

Xuan Zhang, Navid Rajabi, Kevin Duh, and Philipp Koehn. 2023. Machine translation with large language models: Prompting, few-shot learning, and fine-tuning with QLoRA. In *Proceedings of the Eighth Conference on Machine Translation*, pages 468–481, Singapore. Association for Computational Linguistics.

A Appendix

A.1 Detailed Dataset Description

EC30. In Table 5, we list the details of the EC40 dataset. We conducted experiments on EC30, a subset of EC40, where we excluded the data of 10 super low-resource languages, resulting in 30 English-centric language pairs with a total of 61M pairwise data. Each resource group consists of languages from 5 families with multiple writing systems.

A.2 Training Setting

For all of the English-centric training, the learning rate is 5e-4 with 4,000 warmup steps and a *inverse sqrt* decay schedule. All dropout rates and label smoothing are set to 0.1. In the case of EC30 and Europarl-8, the batch size is set as 8,196 tokens, accumulating gradients 20 and 8 times, respectively. Also, data from different language pairs are sampled with a temperature of 5.0 and 2.0, respectively. The same temperature is applied to both BPE building and MMT training periods. We train all models with an early-stopping strategy[10] and evaluate by using the best checkpoint as selected based on the loss on the development set.

For fine-tuning, all parameters are kept the same as those in training, except for 1) we set batch accumulation as 1 in Section 4 as extremely small fine-tuning data is used, and 2) we set patience as 3 for quick experiments.

Note that we use 4 A6000 GPU cards for English-centric training with FP16 optimization, which means the actual batch size is also 4 times bigger. For fine-tuning, we use a single A6000 GPU card.

A.3 Detailed Results on EC30

We report our detailed results in 970 directions (including English-centric and zero-shot directions) on EC30 datasets for both one-tag and two-tag models. The results are measured by 3 widely used metrics, i.e., ChrF++, SacreBLEU, and COMET.

Table 6, Table 7 and Table 8 show the specific performance of the two-tag model in each direction measured by ChrF++, SacreBLEU, and COMET, respectively. In each table, we report the corresponding performance of the baseline, boost-100, and boost-all models. We also report the corresponding results in one-tag setting in Table 9, Table 10, and Table 11, respectively. The results grouped by resource level can be found in Table 1, 12, and 14 for ChrF++, SacreBLEU, and COMET, respectively.

We also report the influence fine-tuning brings on English-centric directions, which can be found in Table 2, 13, and 15 for ChrF++, SacreBLEU, and COMET, respectively.

A.4 More Data or More Directions?

A.4.1 Limited Fine-tuning Direction Set

To further investigate the surprising boosting effects that partial directional data brings, we limit the fine-tuning direction set to fall in a specific family and check the corresponding influence across language families. Here, we limit the fine-tuning set within *Germanic* (including English) and also use NTREX to build pairwise samples to cover all of the possible 42 translation directions.

Figure 5 summarizes the zero-shot performance across language groups. It is easy to see that even when limiting fine-tuning to a specific language family (Boost-Germanic), the overall performance remains comparable to full fine-tuning (Boost-All). More specifically, Boost-Germanic achieves a slight improvement over Boost-All in Germanic directions, meanwhile slightly lagging behind in all other groups, which is also intuitive. However, the gap between the two settings is still small. This finding further demonstrates the insensitivity of the directional data during fine-tuning. Detailed results, including those in other metrics, are provided in Table 18.

A.4.2 Detailed Results: More Data or More Directions?

Table 18 shows the detailed results when fine-tuning with Germanic data and all of the NTREX data.

A.5 Detailed Results: How Close to the Upper Bound?

In Table 19, 20 and 21, we show the detailed results for the baseline, boosted, and upper bound models on the Europarl-8 dataset in the metric of ChrF++, COMET, and SacreBLEU, respectively.

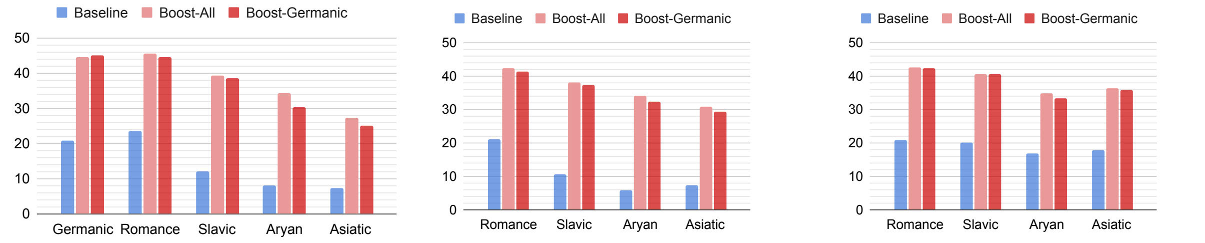
A.6 Number Pairs

The synthetic number pairs and word pairs are illustrated in Figure 6.

[10] Patience is set to {10, 20}, i.e., training stops if performance on the validation set does not improve for the last {10, 20} checkpoints, with 1,000 steps between checkpoints.

Resource	Germanic			Romance			Slavic			Indo-Aryan			Afo-Asiatic		
	ISO	Language	Script	ISO	Language	Script	ISO	Language	Script	ISO	Language	Script	ISO	Language	Script
High (5M)	DE	German	Latin	FR	French	Latin	RU	Russian	Cyrillic	HI	Hindi	Devanagari	AR	Arabic	Arabic
	NL	Dutch	Latin	ES	Spain	Latin	CS	Czech	Latin	BN	Bengali	Bengali	HE	Hebrew	Hebrew
Med (1M)	SV	Swedish	Latin	IT	Italian	Latin	PL	Polish	Latin	KN	Kannada	Devanagari	MT	Maltese	Latin
	DA	Danish	Latin	PT	Portuguese	Latin	BG	Bulgarian	Cyrillic	MR	Marathi	Devanagari	HA	Hausa*	Latin
Low (100K)	AF	Afrikaans	Latin	RO	Romanian	Latin	UK	Ukrainian	Cyrillic	SD	Sindhi	Arabic	TI	Tigrinya	Ethiopic
	LB	Luxembourgish	Latin	OC	Occitan	Latin	SR	Serbian	Latin	GU	Gujarati	Devanagari	AM	Amharic	Ethiopic
eLow (50K)	NO	Norwegian	Latin	AST	Asturian	Latin	BE	Belarusian	Cyrillic	NE	Nepali	Devanagari	KAB	Kabyle*	Latin
	IC	Icelandic	Latin	CA	Catalan	Latin	BS	Bosnian	Latin	UR	Urdu	Arabic	So	Somali	Latin

Table 5: Details of the EC40 dataset. Numbers in the table represent the number of sentences, e.g., 5M denotes exactly 5,000,000 sentences. Two exceptions are Hausa and Kabyle, where the size is 334K and 18K, respectively.



(a) Within Each Family

(b) Out of Germanic

(c) Into Germanic

Figure 5: Zero-shot performance (ChrF++) on EC30. Boost-All means fully fine-tuning, while Boost-Germanic means partially fine-tuning using Germanic languages. (a) shows the average performance evaluated within a specific language group, where both the source and target languages belong. (b) and (c) show the average performance in out-of-Germanic and into-Germanic directions, respectively. Detailed results are provided in Table 18.

ChrF++	en	de	nl	fr	es	ru	cs	hi	bn	ar	he
en	-	57.3/56.6/57.0	52.2/51.8/51.6	63.0/61.4/61.9	49.7/49.1/49.3	48.6/47.1/47.3	50.8/49.2/49.7	54.3/53.4/54.2	50.6/50.6/50.5	46.1/46.1/46.1	51.9/51.3/51.9
de	62.6/60.8/61.2	-	20.8/44.6/44.2	22.0/48.0/46.7	19.9/41.2/40.8	3.0/38.5/37.9	16.2/39.5/38.9	3.0/38.2/37.7	3.3/38.4/37.8	3.6/33.7/32.6	3.1/35.5/34.8
nl	54.9/54.0/54.2	19.0/43.4/42.9	-	21.6/44.0/43.2	19.5/39.4/38.3	2.8/35.1/34.8	16.0/35.3/34.9	2.9/36.0/35.8	3.4/36.8/36.3	3.6/31.5/30.7	3.2/32.8/32.2
fr	61.9/60.6/61.0	19.2/45.0/44.2	20.5/43.3/42.5	-	20.8/44.0/43.7	3.0/39.1/38.3	16.2/38.8/38.1	2.4/39.4/39.2	2.5/38.8/38.0	3.7/35.2/34.8	3.4/37.0/36.5
es	54.1/53.1/53.4	18.7/40.7/40.0	20.3/40.5/39.7	23.0/46.2/45.5	-	3.1/36.1/35.3	16.0/35.0/34.3	2.7/36.0/35.7	2.6/36.2/36.0	3.4/32.3/31.4	3.4/33.0/32.3
ru	54.9/53.2/53.6	18.2/41.1/40.3	19.4/39.3/39.1	21.6/43.6/42.7	19.4/38.7/38.2	-	15.5/36.5/36.1	3.0/34.7/34.8	3.5/35.4/34.9	4.1/31.4/30.5	3.6/32.3/32.1
cs	58.9/57.3/57.5	19.1/43.2/42.3	20.1/41.1/40.4	22.2/45.7/44.2	19.8/40.2/39.2	3.1/38.9/38.2	-	2.7/36.6/36.5	2.5/36.7/35.9	3.6/32.1/31.2	3.4/33.6/33.2
hi	58.5/56.1/57.1	17.7/38.7/37.8	19.2/38.2/37.9	20.2/41.7/40.5	18.6/37.0/36.1	2.8/33.7/33.0	15.1/33.0/32.1	-	4.9/41.2/40.9	2.9/30.9/29.6	2.8/30.8/30.0
bn	54.1/52.0/52.5	17.4/36.5/35.6	18.8/36.6/35.7	19.9/39.1/38.0	18.2/34.8/33.8	2.6/32.3/31.1	14.9/30.8/30.0	4.2/39.6/38.5	-	2.6/28.9/27.8	2.8/29.2/28.0
ar	54.8/52.5/52.7	17.3/37.7/36.8	18.4/36.8/36.2	21.0/41.4/40.6	19.0/36.8/36.1	3.7/33.4/33.1	14.6/31.8/31.4	4.1/32.8/32.7	4.5/33.9/32.6	-	3.8/31.5/31.1
he	60.0/57.2/58.2	17.6/40.1/39.6	19.0/39.3/38.8	20.9/43.6/42.8	18.8/38.3/37.4	3.1/35.8/34.9	15.0/34.7/34.0	3.0/34.5/33.9	3.1/34.9/33.9	3.3/32.2/31.4	-
sv	63.3/61.9/61.8	19.8/46.2/45.3	20.7/43.4/42.9	22.5/48.1/46.9	20.0/40.8/40.0	3.5/37.7/37.4	16.3/38.6/37.6	4.2/37.9/37.9	4.3/38.9/37.7	4.2/33.4/32.3	3.5/35.2/34.5
da	64.0/62.6/62.7	19.9/46.7/45.6	21.1/43.9/43.3	22.4/48.0/47.1	20.2/41.1/40.3	3.4/38.2/38.0	16.5/38.9/38.0	3.6/38.8/38.2	4.1/38.7/37.5	4.4/33.8/32.6	3.7/35.5/34.4
it	56.2/55.1/55.4	19.4/42.1/41.8	20.6/41.3/40.9	23.7/48.8/48.0	21.1/43.5/43.2	3.4/36.6/36.4	16.4/36.2/35.6	3.6/36.7/36.7	4.3/37.0/36.3	4.0/32.8/32.2	3.6/33.6/33.2
pt	64.4/62.8/63.0	18.9/44.0/43.6	20.5/42.5/41.9	23.7/51.7/51.1	21.2/45.3/45.0	3.2/38.5/38.1	16.1/38.2/37.5	3.0/38.3/38.6	3.7/38.8/37.9	3.9/34.2/33.9	3.4/35.9/35.1
pl	51.2/50.0/50.4	18.9/39.2/38.6	19.9/39.1/38.5	22.4/42.3/41.4	20.1/38.4/37.8	3.7/36.1/35.9	16.2/36.0/35.4	3.7/33.9/33.7	4.4/33.9/33.2	4.7/30.0/29.3	3.5/31.1/30.5
bg	61.1/59.5/59.8	18.5/43.7/42.9	19.9/41.8/41.3	21.9/47.3/46.5	19.6/40.7/40.0	4.8/41.7/41.9	15.8/38.6/38.2	3.7/37.4/37.1	3.8/37.5/36.9	4.2/33.4/33.0	3.7/35.4/34.8
kn	50.7/49.2/49.6	17.2/35.4/34.1	18.6/35.6/34.5	20.0/37.2/36.2	18.5/34.1/33.5	3.3/30.3/29.5	14.9/29.3/28.4	7.8/37.2/36.6	7.0/37.6/36.3	3.5/27.5/26.0	3.0/27.3/26.8
mr	52.2/50.4/51.1	17.3/35.6/34.5	18.6/35.9/35.0	19.8/37.6/36.3	18.2/34.4/33.5	3.5/31.2/30.3	14.9/30.2/28.9	10.2/40.0/39.1	7.3/39.2/38.1	3.7/28.1/26.8	2.8/28.6/27.5
mt	70.6/68.2/68.9	19.4/44.9/43.7	20.9/43.1/42.4	23.4/50.8/49.3	20.5/43.2/42.2	2.9/38.4/37.7	16.3/38.7/38.1	3.1/39.8/39.2	3.3/39.1/38.0	3.9/35.3/33.9	3.4/37.0/35.6
ha	31.3/42.1/42.1	10.9/31.7/30.7	11.4/32.2/31.2	14.1/33.6/32.3	11.6/31.4/30.5	2.4/27.3/26.3	9.6/26.9/25.7	1.9/28.7/27.7	2.6/28.9/27.3	2.5/25.7/23.9	2.4/25.5/24.7
af	68.6/65.8/66.8	21.1/44.2/42.7	21.8/43.3/42.3	25.0/45.9/44.2	21.6/39.3/37.9	5.1/36.3/35.2	16.6/36.8/35.2	5.3/37.8/36.4	6.2/37.9/36.2	8.3/31.9/30.0	5.1/34.2/33.0
am	46.3/43.7/44.2	19.4/32.0/31.2	18.9/32.1/31.3	23.4/33.3/32.3	20.9/31.1/30.5	9.6/27.1/26.4	15.2/26.0/25.5	9.7/29.8/29.1	12.0/30.3/29.5	12.3/24.6/23.6	7.2/25.4/25.0
gu	51.0/49.8/49.8	19.2/34.3/33.2	18.9/34.3/33.2	22.3/36.0/34.7	19.7/32.7/32.0	8.3/29.4/28.3	15.0/28.5/27.4	17.3/39.1/38.8	13.3/37.7/36.8	8.6/26.7/24.7	4.7/27.1/26.1
lb	47.2/49.8/47.9	23.5/43.7/41.3	20.4/38.3/36.1	21.8/39.1/35.7	19.9/35.2/33.0	6.4/31.1/28.6	15.9/31.3/28.6	5.5/31.1/29.0	5.6/32.0/29.5	6.1/26.2/23.1	4.1/28.3/25.7
oc	60.4/59.5/59.7	24.6/40.8/40.0	22.4/39.8/39.2	33.5/49.9/49.3	27.2/42.0/41.9	10.0/34.2/33.8	17.2/34.8/33.6	8.2/35.1/34.6	8.5/35.0/34.1	11.2/30.7/29.8	7.4/32.4/32.0
ro	58.1/56.9/57.0	22.4/42.5/41.5	22.1/40.9/40.1	29.2/48.7/47.5	24.6/41.7/41.3	6.6/37.0/36.6	17.6/36.6/36.2	9.0/36.2/35.6	8.4/36.4/35.1	9.1/32.6/31.5	5.0/34.1/33.6
sd	16.3/31.1/27.0	7.5/26.0/23.8	6.5/25.9/23.9	9.3/25.8/23.1	8.3/25.4/23.3	4.0/21.6/19.5	5.3/20.3/18.1	1.8/26.2/23.3	2.6/26.0/21.9	3.1/19.2/15.5	2.3/19.7/17.3
sr	57.1/54.8/55.2	21.8/40.6/39.3	20.0/38.9/37.8	26.5/43.3/41.6	22.0/38.1/37.2	9.2/37.1/36.3	16.8/36.5/35.4	9.1/34.1/33.4	9.7/34.4/32.9	10.2/30.6/29.7	5.7/32.1/31.0
ti	39.9/38.9/38.8	16.8/30.2/29.2	17.1/30.4/29.6	19.2/30.5/29.7	17.2/29.4/28.6	5.2/25.7/25.1	14.3/24.4/23.8	8.6/27.8/26.9	8.3/28.7/27.1	6.8/23.3/22.2	4.2/23.8/22.8
uk	53.5/52.6/52.5	22.2/39.3/38.2	20.2/38.0/37.4	26.7/42.2/41.4	22.5/37.1/36.7	17.6/41.1/40.9	17.3/34.8/34.3	9.2/33.4/33.1	9.6/33.3/32.4	11.5/29.5/28.4	6.7/31.0/30.5

ChrF++	sv	da	it	pt	pl	bg	kn	mr	mt
en	60.3/58.2/58.7	60.3/58.7/59.4	52.1/51.5/51.5	59.6/57.9/58.6	43.6/43.4/43.6	61.3/59.6/60.3	51.5/50.6/51.0	44.8/46.5/45.8	65.8/64.7/65.6
de	21.1/46.5/45.7	21.7/46.8/45.9	21.2/41.7/40.8	20.7/43.7/43.0	16.6/35.8/35.2	3.7/45.5/44.8	7.6/38.3/39.1	8.2/36.3/35.2	17.2/42.7/41.2
nl	20.5/41.8/41.0	21.4/42.3/41.6	20.7/39.2/38.3	20.2/40.7/40.2	16.3/33.6/33.3	3.4/41.4/40.8	8.4/36.5/37.3	8.4/34.9/33.6	16.8/39.7/38.7
fr	21.0/44.6/44.1	21.4/45.2/44.5	21.9/45.9/45.4	21.5/47.9/47.5	16.3/36.5/35.8	3.6/46.3/45.9	6.6/38.9/39.9	6.9/37.2/35.9	17.7/46.1/44.6
es	20.4/40.6/40.0	21.1/41.1/40.5	22.0/42.6/42.1	21.7/45.1/44.8	16.3/34.7/33.8	3.6/42.1/41.5	7.5/37.3/37.4	8.0/35.0/33.6	17.4/41.2/40.2
ru	19.6/39.6/39.1	20.3/40.0/39.5	20.5/38.7/38.1	20.0/40.3/39.6	16.0/34.8/34.6	6.5/45.3/45.0	8.6/35.1/36.2	8.4/33.8/32.6	16.3/38.5/37.2
cs	20.6/42.4/41.3	21.1/42.4/41.8	21.1/40.3/39.3	20.7/42.2/41.4	17.3/36.5/35.7	4.4/44.6/43.9	6.2/36.3/37.2	5.9/35.5/34.1	17.4/41.1/39.5
hi	19.3/38.4/37.6	20.0/38.8/38.2	19.7/36.6/35.6	19.4/38.8/38.2	15.4/30.8/30.1	3.5/39.5/38.5	8.8/40.8/41.1	9.5/40.4/39.3	16.4/38.3/36.8
bn	19.2/36.2/35.2	19.7/36.7/35.9	19.2/34.8/33.9	18.9/36.5/35.7	14.9/29.5/28.5	3.2/37.3/36.2	8.2/38.8/38.9	7.9/37.7/36.5	16.0/35.2/33.8
ar	18.7/37.3/36.4	19.3/37.8/37.2	19.7/36.7/35.7	19.3/38.3/37.7	14.9/30.4/29.7	4.3/39.6/38.9	10.1/34.4/34.6	9.6/32.4/30.9	16.0/37.7/36.3
he	19.2/40.1/39.1	19.9/40.3/39.6	19.7/37.9/37.2	19.7/40.3/39.9	15.2/32.0/31.3	3.9/41.6/40.9	6.8/34.8/35.0	7.8/33.6/31.9	16.3/39.3/37.4
sv	-	22.6/50.6/50.3	21.2/41.4/40.5	20.8/43.9/43.3	16.8/35.3/34.5	3.8/45.1/44.2	8.6/37.9/37.8	8.3/36.5/35.2	17.6/43.1/41.8
da	22.6/50.8/50.6	-	21.3/41.5/40.4	21.1/44.1/43.2	17.1/35.4/34.8	4.0/45.3/44.3	8.9/38.4/38.8	8.7/36.7/35.6	17.6/43.8/42.3
it	20.9/41.5/40.5	21.3/41.7/41.2	-	21.6/44.7/44.5	17.0/34.8/34.3	4.0/43.2/42.9	8.8/37.3/37.3	9.0/35.1/33.9	17.8/42.7/41.5
pt	21.0/44.4/44.0	21.4/45.0/44.5	21.9/44.6/44.1	-	16.5/35.2/34.6	3.7/46.1/45.1	7.6/38.0/39.0	7.4/36.4/35.6	17.8/46.8/45.7
pl	20.2/38.1/37.7	20.9/38.7/37.8	21.3/38.4/37.8	20.3/38.9/38.3	-	5.1/41.4/41.0	9.5/34.3/34.8	9.3/33.0/32.1	16.9/37.6/36.5
bg	20.3/42.8/42.3	21.0/43.1/42.7	20.9/41.0/40.2	20.6/43.8/43.1	16.5/36.0/35.7	-	9.2/36.8/37.9	9.1/35.7/34.3	16.9/42.2/41.0
kn	19.0/34.6/33.6	19.7/34.9/34.3	19.3/33.2/32.4	19.0/35.2/34.5	15.5/28.2/27.2	4.2/35.5/34.5	-	13.2/36.9/35.1	15.7/34.2/32.9
mr	19.1/35.6/34.5	19.5/36.0/35.1	19.3/33.6/32.4	18.7/35.5/34.7	15.5/29.0/27.8	4.6/36.2/34.7	14.2/39.3/39.1	-	15.3/34.4/33.0
mt	21.4/45.3/44.0	22.1/45.7/44.5	22.4/44.3/43.1	21.8/48.6/47.4	16.6/35.9/34.9	4.1/46.3/45.0	8.5/39.3/39.6	8.5/37.6/35.6	-
ha	14.1/31.5/30.3	14.4/31.8/31.0	15.3/30.5/29.3	12.9/31.9/31.1	11.5/26.4/26.0	4.1/31.9/30.5	8.4/30.5/30.1	5.9/29.2/27.1	11.4/31.1/29.7
af	23.2/44.7/43.5	23.6/45.2/44.2	23.6/39.5/38.0	23.2/43.4/41.8	19.6/33.2/32.2	9.6/43.1/41.2	12.1/37.7/37.3	12.6/36.1/34.1	18.4/43.3/40.8
am	20.8/30.9/29.9	20.8/31.4/30.7	22.0/30.2/29.6	20.4/31.3/30.8	16.6/25.5/25.1	17.6/32.0/31.0	19.3/31.4/32.2	16.2/30.0/28.7	15.4/30.4/28.9
gu	20.3/33.7/32.7	20.8/34.0/33.0	21.2/32.1/30.7	19.7/33.8/32.8	18.0/27.2/26.1	11.2/34.4/32.8	22.1/37.9/37.9	20.6/38.0/36.2	15.5/32.7/31.4
lb	20.6/38.4/35.3	21.4/39.1/35.7	21.5/34.8/32.4	20.0/36.9/34.1	17.2/29.2/27.4	9.3/36.7/33.5	10.3/32.0/30.8	10.3/30.6/27.4	15.9/35.2/32.0
oc	24.5/40.8/40.0	24.3/41.5/40.7	29.9/42.1/41.3	28.9/46.3/45.8	21.3/32.3/31.6	17.2/41.4/40.5	16.1/34.7/34.9	14.7/33.3/31.7	20.3/43.1/41.5
ro	23.7/42.0/41.2	23.6/42.4/41.3	26.8/42.7/41.6	24.5/45.0/44.2	20.2/34.9/34.4	8.7/44.0/43.3	17.3/36.0/36.6	14.8/34.5/33.2	19.9/43.4/41.7
sd	8.3/24.8/22.1	10.6/25.0/22.4	10.0/24.4/22.0	8.0/24.9/22.3	7.7/21.2/19.2	4.5/25.0/22.2	8.0/24.7/23.9	8.5/26.1/21.2	7.6/22.7/19.3
sr	21.8/40.1/38.8	23.3/40.4/39.3	24.8/38.0/36.8	22.4/40.3/39.0	20.5/33.6/32.7	13.9/44.3/43.4	17.8/34.5/34.1	16.6/33.3/31.7	18.1/38.9/37.4
ti	17.6/28.9/28.0	18.3/29.2/29.0	18.5/28.3/27.5	17.5/29.3/28.6	14.9/24.1/23.7	7.5/29.8/29.2	14.5/28.7/29.9	14.3/28.2/27.1	15.4/27.8/26.7
uk	22.7/38.3/37.6	21.8/38.7/38.4	23.3/37.0/36.0	22.6/38.6/38.3	20.8/33.6/32.9	19.3/43.1/42.8	15.6/32.6/34.2	15.0/32.7/31.0	17.9/36.6/35.4

ChrF++	ha	af	am	gu	lb	oc	ro	sd	sr	ti	uk
en	40.1/43.7/42.7	62.8/62.0/62.7	29.3/31.2/30.9	46.5/47.6/47.0	42.6/45.8/44.7	52.0/52.9/53.2	51.6/52.5/52.1	18.8/23.3/22.4	51.3/49.7/49.6	22.4/23.7/22.8	47.3/45.9/46.5
de	14.5/31.8/29.9	20.8/45.8/44.6	1.0/19.9/18.1	4.4/34.5/33.2	19.3/41.3/39.9	21.0/39.3/38.8	20.3/41.2/39.2	1.5/18.5/16.2	16.5/37.7/36.5	1.7/17.8/15.5	2.5/35.5/34.3
nl	14.4/31.4/29.6	20.6/43.9/43.2	1.0/18.9/17.3	4.7/32.5/31.2	18.2/35.2/33.9	20.7/37.3/36.6	19.8/37.8/36.1	1.5/17.8/16.1	16.2/34.7/33.4	1.9/16.8/14.8	2.5/32.6/31.6
fr	14.5/32.7/30.4	20.7/45.1/44.3	1.0/20.3/18.3	3.5/35.4/34.6	18.6/36.3/34.4	22.3/45.2/45.1	20.9/44.4/43.0	1.4/19.5/17.0	16.4/38.5/37.0	1.6/17.5/15.4	2.5/36.3/35.0
es	14.4/31.2/29.1	20.2/40.7/40.1	1.0/18.4/16.2	4.0/32.6/31.4	18.1/33.4/31.9	21.6/40.2/40.0	20.5/40.2/38.4	1.5/18.2/15.6	16.5/35.4/34.3	1.6/16.4/14.3	2.6/33.0/32.0
ru	13.8/30.1/28.3	19.7/39.5/38.8	1.1/17.4/15.7	4.8/31.8/30.6	17.5/31.9/30.4	20.3/36.2/35.3	19.5/37.9/36.1	1.6/17.4/14.9	16.6/36.4/35.3	2.0/15.9/13.9	3.7/39.6/38.4
cs	14.4/31.8/29.3	20.4/41.9/40.9	1.1/18.6/16.7	3.6/33.1/31.4	18.3/34.3/32.0	21.0/37.8/37.0	20.3/39.9/37.8	1.5/17.9/15.3	16.8/38.1/36.5	1.6/16.5/14.2	2.7/35.7/34.4
hi	14.0/30.5/28.9	19.3/40.2/39.0	1.2/19.3/16.5	5.7/39.3/37.9	17.4/30.6/29.0	19.4/34.4/33.8	19.1/36.0/34.1	1.5/20.5/17.0	15.8/32.6/31.7	1.7/16.8/14.1	2.5/30.7/29.2
bn	13.7/29.1/27.3	18.9/37.6/36.2	1.1/18.3/15.3	4.8/36.4/34.7	16.9/29.5/27.3	19.0/32.6/31.6	18.7/33.9/32.3	1.3/19.0/15.4	15.6/30.9/29.5	1.6/16.2/13.6	2.5/28.9/27.3
ar	13.5/28.9/27.0	19.0/38.0/37.1	1.1/16.8/15.0	4.9/30.0/28.5	16.7/29.1/27.7	19.8/34.5/33.9	18.8/35.8/33.7	1.6/17.0/15.2	15.8/32.6/31.1	2.0/15.4/12.9	2.6/30.3/29.4
he	13.7/30.6/28.6	19.4/40.9/39.8	1.0/17.9/15.9	3.7/31.4/29.7	17.1/31.1/29.4	19.9/36.2/35.2	19.1/37.8/35.8	1.5/17.1/15.8	16.0/34.5/32.7	1.7/16.2/13.5	2.7/32.4/31.0
sv	14.4/32.7/30.1	20.8/46.4/45.4	1.0/19.6/17.6	4.4/34.8/33.1	18.6/37.2/35.0	21.2/39.6/39.0	20.2/41.2/39.6	1.5/18.9/16.9	16.4/37.6/36.2	1.6/17.0/14.8	2.6/35.0/33.9
da	14.6/33.5/31.0	21.0/47.0/45.8	1.0/20.1/18.3	4.3/34.7/33.1	18.7/37.6/35.8	21.3/40.6/39.8	20.4/41.8/39.4	1.6/19.0/16.8	16.4/38.0/36.4	1.7/17.4/15.0	2.7/35.1/34.1
it	14.5/30.5/28.3	20.5/41.6/40.7	1.0/18.9/16.8	4.5/33.3/31.8	18.6/34.0/32.6	21.8/41.1/40.4	21.0/41.2/39.7	1.5/18.4/16.1	16.5/35.9/34.7	1.5/16.7/14.3	2.7/34.0/32.8
pt	14.6/33.1/30.5	20.8/45.0/44.2	1.0/19.9/18.3	4.0/34.8/33.6	18.5/35.6/34.1	22.2/44.7/44.4	20.8/43.7/42.0	1.4/19.7/17.0	16.4/38.0/36.5	1.5/17.0/14.9	2.7/35.7/34.2
pl	14.2/29.7/27.9	19.8/38.5/37.7	1.0/17.1/14.9	5.1/30.5/29.5	18.0/31.3/29.8	20.9/35.2/34.9	20.1/36.8/35.4	1.5/16.9/15.3	16.8/34.4/33.3	2.1/15.8/13.5	2.9/33.5/32.4
bg	14.0/31.5/30.3	20.1/43.0/42.0	1.1/19.3/17.7	4.8/33.6/32.4	18.1/34.5/32.9	21.0/39.6/38.8	19.9/41.3/39.1	1.6/18.2/16.7	17.2/39.4/38.0	2.1/16.9/14.7	3.3/38.1/36.9
kn	13.6/28.2/26.7	18.8/36.0/34.7	1.2/18.1/15.5	8.3/34.2/32.5	16.4/28.1/26.8	18.8/31.8/30.8	18.4/32.7/30.6	1.6/18.5/15.9	15.8/29.6/28.5	2.3/15.8/13.1	3.1/27.2/25.8
mr	13.3/28.3/25.3	18.7/37.2/35.4	1.4/17.8/15.1	8.5/37.4/35.7	16.2/28.4/26.6	18.5/31.8/30.5	18.6/32.9/31.2	1.7/18.6/15.1	15.7/30.0/28.5	2.3/15.9/13.0	2.9/28.3/26.7
mt	15.1/35.0/32.3	21.5/46.9/45.7	1.2/20.7/18.1	4.4/35.6/34.3	18.8/36.5/34.2	22.3/43.0/42.0	20.9/43.6/41.1	1.7/18.5/15.9	16.7/38.6/36.7	2.2/17.7/14.5	2.6/35.5/33.9
ha	-	13.9/33.4/32.1	1.1/14.6/11.7	4.7/26.9/24.4	11.1/26.1/24.2	10.4/29.9/28.7	14.2/30.3/28.0	2.4/15.7/11.0	12.7/27.3/25.6	2.1/13.9/10.8	2.0/25.0/23.5
af	15.1/31.8/29.6	-	1.4/20.0/17.0	7.4/34.2/32.0	18.6/36.7/34.5	22.1/38.9/38.0	20.7/39.4/36.8	2.6/18.9/16.6	19.3/35.5/33.2	3.4/16.8/13.8	3.6/33.0/31.2
am	13.7/25.7/23.1	19.1/32.4/31.5	-	11.7/27.7/25.9	15.3/24.6/23.9	18.3/28.0/27.1	18.6/29.0/27.7	5.1/14.9/12.7	17.0/26.8/25.3	9.4/17.8/15.1	6.4/24.9/23.3
gu	13.4/27.3/24.8	18.8/35.2/34.0	2.2/16.4/14.2	-	15.9/27.5/25.9	18.7/30.2/29.2	18.4/31.6/29.2	2.5/19.4/15.9	17.1/28.4/26.8	4.3/14.9/11.8	5.4/26.8/25.2
lb	14.5/28.0/24.5	20.0/40.3/36.5	1.3/15.3/12.2	6.4/29.0/25.7	-	18.9/33.8/31.4	18.8/34.3/30.4	3.2/15.9/12.7	17.4/30.8/27.9	3.0/14.1/10.3	4.1/28.1/25.1
oc	15.5/30.2/28.5	22.1/42.2/40.7	2.0/17.0/15.4	9.3/31.6/29.9	18.1/33.5/31.8	-	22.9/41.0/38.9	3.5/17.9/16.1	20.3/34.4/32.9	4.3/15.3/12.8	5.4/31.3/30.0
ro	14.5/31.6/29.0	21.4/42.2/40.9	1.1/18.0/16.4	7.4/32.8/31.2	19.0/34.0/32.4	22.9/41.1/40.6	-	1.9/18.4/16.7	18.0/36.7/35.0	3.4/16.2/14.1	3.3/34.0/32.5
sd	12.3/21.6/17.5	8.8/27.1/23.9	0.9/11.1/6.1	5.2/25.2/19.9	6.9/20.3/16.9	6.3/23.4/20.5	9.1/23.8/20.1	-	10.2/22.0/19.9	1.7/10.9/6.5	1.8/20.1/17.9
sr	14.0/30.2/28.4	21.0/40.2/38.8	1.4/17.4/15.2	8.2/31.2/29.2	17.9/32.0/29.7	22.0/35.6/34.6	20.2/37.9/35.5	2.5/17.1/16.0	-	4.1/15.5/13.1	4.0/34.2/32.1
ti	13.2/24.8/22.9	18.1/30.6/29.6	2.2/17.7/16.3	6.3/25.8/24.3	15.4/23.7/23.1	16.9/26.2/25.9	17.0/27.4/25.8	2.7/14.7/13.7	15.4/25.1/24.0	-	3.7/23.4/22.1
uk	13.9/28.7/26.2	20.1/38.6/37.4	1.3/16.3/14.5	7.5/30.1/28.4	17.4/30.7/28.8	21.0/34.6/33.6	20.3/36.1/33.9	2.3/17.0/14.9	18.6/34.6/33.1	4.2/15.0/12.3	-

Table 6: The ChrF++ performance on EC30 for the baseline, boost-100, and boost-all models in the two-tag fashion, respectively. 970 directions of results are shown, including both English-centric and zero-shot ones. Overall, in all 870 zero-shot directions, the boost-100 model achieves better performance compared to the baseline. Moreover, in 845 out of 870 zero-shot directions, the gain exceeds 10.0.

SacreBLEU	en	de	nl	fr	es	ru	cs	hi	bn	ar	he
en	-	31.0/30.1/30.1	23.4/22.7/22.9	40.6/39.5/39.0	23.1/22.9/23.2	22.4/21.1/21.6	25.0/23.5/23.4	38.6/38.5/37.5	32.0/32.1/31.7	19.2/18.7/19.4	24.6/23.7/22.2
de	37.6/36.3/35.3	-	2.5/16.4/16.5	2.9/21.7/23.4	2.2/15.5/16.5	1.2/13.1/13.8	1.9/14.9/15.5	1.4/20.9/21.4	2.0/18.9/19.9	0.8/9.4/10.2	0.8/10.5/10.5
nl	28.2/27.2/27.5	2.6/15.5/16.0	-	3.1/17.8/18.9	2.0/13.8/14.5	0.9/10.0/10.5	1.9/11.6/12.1	1.3/19.1/19.5	2.0/16.8/18.2	0.6/7.2/8.2	0.9/8.5/8.6
fr	36.3/35.1/35.3	2.5/17.6/18.3	2.4/14.4/14.9	-	2.2/17.9/18.3	0.9/13.2/14.3	1.7/13.9/14.2	0.8/21.7/22.2	1.2/18.7/19.8	0.7/10.5/11.2	0.9/11.6/11.2
es	25.1/24.1/24.2	1.9/12.4/12.8	1.9/11.3/11.7	3.2/19.1/19.7	-	0.9/10.2/10.9	1.4/10.3/11.0	1.0/17.8/18.5	1.2/16.2/17.5	0.6/7.6/8.0	0.8/7.8/7.8
ru	28.1/26.8/26.5	1.8/14.1/14.5	1.5/11.2/11.5	2.5/18.2/18.8	1.8/13.7/14.4	-	1.3/12.2/12.7	1.0/17.4/18.1	1.4/15.9/17.1	0.7/7.9/8.3	0.8/8.4/8.5
cs	33.0/31.6/31.3	2.6/16.2/16.8	2.2/13.2/13.3	3.0/20.1/21.2	2.1/14.0/15.2	1.1/13.9/14.1	-	1.0/20.3/19.8	1.2/17.4/18.6	0.8/8.1/9.1	0.9/9.5/9.3
hi	33.6/31.6/31.0	1.8/12.2/13.0	1.9/10.7/10.9	2.0/16.0/17.0	1.6/11.6/12.3	0.8/9.6/10.5	1.4/10.3/10.6	-	3.1/22.6/23.2	0.6/7.7/8.6	0.7/7.5/7.9
bn	28.1/26.3/25.8	1.9/10.5/10.9	1.9/9.7/10.3	2.0/14.0/14.9	1.5/10.6/11.6	0.8/8.6/8.9	1.3/8.2/8.8	2.3/22.6/22.5	-	0.5/6.4/6.9	0.7/6.9/7.2
ar	28.3/26.7/27.0	1.6/11.8/12.2	1.3/9.8/10.6	2.4/16.5/17.6	1.7/11.6/12.4	1.0/9.5/10.3	1.1/9.6/9.8	1.4/16.6/16.8	1.7/14.2/15.2	-	0.7/8.5/8.1
he	35.9/33.7/32.7	1.4/14.2/14.9	1.5/12.0/11.8	2.3/19.6/21.1	1.5/13.5/13.9	0.9/11.4/12.0	1.2/11.4/11.9	1.3/18.2/18.8	1.7/16.3/17.5	0.6/8.6/9.2	-
sv	39.3/38.2/38.1	2.9/19.7/20.4	2.5/15.7/16.4	3.3/23.1/24.0	2.2/15.5/16.0	1.0/13.6/13.8	1.9/14.0/14.7	2.2/20.7/21.7	2.6/18.6/19.5	0.8/9.1/10.4	0.9/10.9/11.4
da	39.5/38.3/38.4	2.9/19.8/19.9	2.6/16.4/16.4	3.2/22.9/24.4	2.3/15.7/16.6	1.1/12.9/13.3	2.0/14.7/15.4	1.6/21.6/21.5	2.4/18.6/19.6	0.8/9.3/10.6	0.9/10.8/10.9
it	27.4/26.8/27.0	2.6/14.3/14.7	2.2/12.4/13.0	3.7/22.6/23.4	2.4/16.9/17.6	1.0/11.1/11.7	1.8/11.3/12.2	1.4/19.2/19.4	2.3/16.8/18.1	0.7/7.7/8.5	0.8/8.7/8.5
pt	39.5/38.3/38.3	2.4/17.8/18.0	2.3/14.2/14.9	3.9/27.3/28.2	2.5/19.2/19.9	1.1/13.1/13.6	1.7/14.1/14.5	1.4/21.9/22.4	2.1/19.0/20.0	0.9/9.8/10.9	0.8/11.3/11.3
pl	23.2/22.3/22.8	2.4/12.2/12.9	2.1/11.2/11.6	3.1/16.9/17.9	2.2/13.2/13.6	1.1/11.7/11.7	2.0/12.2/12.7	1.2/17.5/17.3	2.2/15.0/15.7	0.8/6.5/7.5	0.7/7.3/7.5
bg	34.9/33.2/33.3	2.0/16.1/17.1	1.9/13.1/13.7	2.5/21.7/22.5	1.8/14.9/15.4	1.5/15.9/16.0	1.5/13.8/14.2	1.5/20.2/20.6	2.1/18.5/19.1	0.8/9.3/10.1	0.8/10.1/10.4
kn	24.5/23.6/23.4	1.7/9.6/9.7	1.7/9.1/9.0	2.0/12.4/13.4	1.5/9.8/10.6	0.8/7.1/7.5	1.4/7.7/8.2	4.4/20.8/20.7	3.9/18.9/19.3	0.7/5.4/6.2	0.7/6.0/5.9
mr	26.1/24.4/24.5	1.9/10.3/10.8	1.7/9.5/9.6	2.3/13.0/14.0	1.6/10.1/10.6	1.0/8.1/8.4	1.4/8.1/8.9	5.3/23.8/23.5	4.4/20.5/21.2	0.6/5.9/6.6	0.7/6.4/6.5
mt	48.8/46.3/46.1	2.6/17.9/18.8	2.5/14.4/15.1	3.7/25.7/27.0	2.3/17.1/18.1	1.1/13.2/13.2	1.9/14.3/15.4	1.6/23.3/23.6	1.9/20.0/20.4	0.9/10.5/11.7	1.0/11.7/11.8
ha	10.4/19.5/19.3	0.5/8.4/9.1	0.5/7.4/8.1	1.2/11.1/12.2	0.6/8.4/9.4	0.2/6.4/7.3	0.5/6.9/7.3	0.4/12.1/12.6	0.4/10.2/11.5	0.2/5.0/5.8	0.1/5.3/5.3
af	48.7/45.9/45.0	3.9/17.8/18.5	3.1/15.3/16.0	5.2/21.5/23.2	3.1/14.1/15.2	1.7/12.3/12.9	2.1/13.4/13.7	2.1/21.0/21.4	3.3/18.3/18.9	1.8/9.3/10.2	1.4/10.5/10.7
am	21.6/19.2/19.5	2.6/8.1/8.1	2.1/7.4/7.7	4.1/9.8/10.8	2.8/7.9/8.3	1.8/5.9/6.5	1.5/6.0/6.3	2.9/13.6/13.4	3.9/12.4/12.9	1.7/4.5/4.9	1.2/4.7/5.0
gu	24.9/23.6/23.6	2.5/9.2/9.5	1.8/8.1/8.5	3.4/11.9/13.3	2.4/9.2/10.0	1.8/6.9/7.9	1.3/7.5/7.6	8.8/22.2/22.2	6.8/18.7/19.8	1.1/4.7/5.6	0.9/5.5/6.1
lb	21.6/21.9/24.6	4.4/15.6/17.6	2.4/11.2/12.4	3.2/13.3/15.9	2.3/10.1/11.8	1.2/7.0/8.6	1.8/8.5/9.7	1.7/13.7/15.1	1.8/11.7/13.9	1.0/4.6/5.6	0.6/5.4/6.3
oc	36.0/34.4/35.0	4.8/14.6/14.8	3.2/12.1/12.4	10.8/25.0/26.2	6.0/16.1/16.7	2.2/10.3/10.7	2.2/11.3/11.4	2.5/18.5/18.6	3.1/15.8/17.1	2.0/7.6/8.4	1.3/8.9/9.0
ro	31.1/29.7/31.2	3.9/15.2/16.0	2.9/12.6/12.8	7.9/23.8/24.7	4.6/16.2/16.4	1.9/12.0/12.9	2.5/13.0/13.8	4.4/19.9/19.9	4.2/17.1/18.1	1.7/8.4/9.3	1.1/9.8/9.8
sd	1.6/5.7/8.6	0.2/2.9/4.3	0.1/2.5/3.9	0.2/3.1/5.2	0.2/3.1/4.4	0.1/2.5/3.6	0.1/2.1/3.6	0.2/6.9/10.0	0.2/5.6/8.0	0.1/1.7/2.7	0.0/1.9/2.6
sr	31.4/29.0/29.0	3.4/13.6/14.9	1.8/11.2/11.5	5.3/18.3/19.4	3.2/13.2/14.1	2.4/12.4/13.1	1.7/13.0/13.4	4.1/17.9/18.0	4.8/15.4/15.9	1.8/7.7/8.5	1.3/8.5/8.8
ti	14.5/14.0/14.1	1.6/6.0/6.3	1.2/5.7/6.1	2.0/8.0/8.6	1.3/6.3/6.8	1.1/5.0/5.4	1.1/4.9/4.9	3.6/12.0/11.6	3.8/11.4/11.7	0.9/3.7/4.1	0.8/3.9/3.9
uk	26.8/26.2/27.0	3.3/12.8/13.4	1.8/10.9/11.0	5.8/17.0/17.9	3.6/12.3/13.1	5.3/16.3/16.1	2.0/11.5/12.3	3.4/16.4/16.8	4.0/14.5/15.0	1.9/6.7/7.6	1.2/7.9/7.9

SacreBLEU	sv	da	it	pt	pl	bg	kn	mr	mt
en	35.8/33.5/33.6	35.9/34.7/34.8	24.4/23.7/23.7	34.4/33.2/33.1	16.2/15.9/16.3	35.7/34.1/33.9	33.6/34.1/30.0	26.6/27.4/27.5	37.9/37.7/37.6
de	2.4/20.1/20.8	2.9/21.7/22.5	2.5/15.8/16.1	2.4/17.6/18.6	2.1/11.0/11.9	1.7/20.3/20.1	4.3/21.1/21.0	3.8/16.1/17.5	2.0/15.1/16.3
nl	2.3/15.2/15.8	2.8/16.0/17.2	2.3/12.9/13.4	2.4/14.7/15.4	2.1/9.6/9.6	1.4/16.0/16.4	5.0/19.4/19.8	3.4/15.2/16.3	1.8/12.3/13.7
fr	2.4/18.2/19.2	2.6/19.0/20.2	2.6/19.1/19.3	2.6/21.7/22.5	1.8/11.0/11.3	1.5/19.6/19.9	3.6/21.4/21.3	3.3/16.3/17.8	2.1/17.6/18.8
es	1.8/12.6/13.7	2.1/14.4/15.5	2.4/15.0/15.7	2.3/17.0/17.6	1.7/8.8/9.3	1.3/14.9/15.3	3.9/18.7/19.4	3.4/14.0/15.1	1.7/12.3/13.1
ru	1.5/13.6/14.6	1.8/15.1/16.2	1.9/13.2/13.5	1.8/14.0/14.7	1.5/10.2/10.4	2.2/19.1/19.0	4.0/17.6/17.6	3.1/14.0/15.3	1.3/11.2/12.3
cs	2.2/16.6/17.5	2.6/17.7/18.9	2.4/13.8/14.2	2.3/16.3/17.0	2.4/11.9/12.2	1.8/19.1/18.8	2.9/20.5/20.6	2.2/15.4/16.5	1.9/13.4/14.6
hi	1.8/12.6/13.8	2.0/13.8/15.0	1.8/11.1/11.6	1.8/13.3/13.9	1.5/7.1/7.6	1.2/14.9/15.4	5.6/24.8/24.1	5.4/21.3/22.0	1.7/11.4/13.2
bn	2.0/11.1/12.1	2.1/12.6/13.6	1.6/9.7/10.3	1.7/11.3/12.1	1.4/6.8/7.2	1.1/12.8/13.0	5.1/22.7/22.4	4.2/18.9/19.3	1.5/9.3/10.3
ar	1.4/12.0/13.4	1.5/13.4/14.4	1.6/11.1/11.5	1.6/12.8/13.7	1.3/7.3/7.3	1.4/14.6/15.2	4.5/16.6/16.9	3.3/12.9/13.8	1.1/11.8/12.5
he	1.3/15.2/15.9	1.6/16.0/17.1	1.4/12.1/12.9	1.5/15.3/15.7	1.2/8.4/8.9	1.4/17.3/17.2	3.7/18.3/17.2	3.4/14.6/15.1	1.3/13.1/14.5
sv	-	3.4/26.2/26.7	2.4/15.3/15.8	2.6/19.1/19.7	2.0/10.8/11.2	1.4/19.7/20.5	4.8/20.6/21.0	4.2/16.1/17.5	1.9/16.4/17.9
da	3.2/25.3/25.4	-	2.5/15.6/16.2	2.6/18.8/19.8	2.4/10.8/11.2	1.6/19.9/20.3	5.4/21.0/20.9	4.3/16.4/17.5	2.0/15.8/17.4
it	2.3/13.8/14.9	2.5/15.0/16.2	-	2.7/17.5/18.2	2.0/9.6/10.4	1.4/16.8/16.9	5.1/19.3/19.3	3.7/14.9/15.8	2.0/13.9/15.0
pt	2.4/18.2/19.7	2.5/20.2/21.8	2.5/17.7/17.8	-	1.9/10.6/11.0	1.5/20.2/21.0	4.2/21.3/21.5	3.7/16.1/17.6	2.1/18.8/20.2
pl	2.0/12.3/12.9	2.4/13.5/14.6	2.5/12.3/13.1	2.2/13.0/13.5	-	1.9/15.9/16.0	4.5/17.7/17.9	3.4/13.9/14.6	1.8/10.8/11.8
bg	1.9/16.4/17.4	2.3/18.1/18.7	2.0/14.4/15.1	2.1/17.0/17.7	1.8/10.6/11.4	-	5.3/20.5/19.7	3.9/16.2/17.0	1.6/13.5/14.6
kn	1.7/9.7/10.4	2.0/10.6/11.3	1.7/8.6/9.0	1.7/10.3/10.7	1.5/6.2/6.2	1.3/11.1/12.0	-	6.2/18.7/18.3	1.4/8.4/9.4
mr	2.0/11.0/11.4	2.1/11.7/13.0	1.9/9.0/9.7	1.8/10.5/11.4	1.6/6.5/7.0	1.4/12.3/12.5	7.6/23.1/22.8	-	1.4/8.7/10.1
mt	2.4/18.8/20.2	3.0/20.4/21.9	3.1/17.3/17.8	3.0/22.4/22.7	2.1/11.4/11.7	1.9/20.9/21.0	5.1/22.4/22.4	3.6/17.4/18.2	-
ha	1.1/9.0/9.5	1.0/9.9/10.8	1.5/7.4/8.1	0.8/9.2/10.0	0.8/6.1/6.5	0.4/10.2/10.9	1.4/13.6/14.8	0.6/9.4/11.1	0.9/8.5/9.5
af	3.9/19.4/20.3	4.1/20.3/22.1	4.1/13.6/14.7	4.2/18.6/19.3	3.2/9.7/10.4	3.9/18.7/18.9	5.4/20.3/19.7	4.4/15.9/17.3	2.7/17.1/19.0
am	2.5/7.4/8.3	2.8/8.4/9.3	3.1/6.7/7.3	2.6/8.5/8.9	1.7/4.5/5.0	3.9/9.9/10.0	7.1/15.6/14.6	4.1/11.2/11.9	1.5/6.9/8.1
gu	2.3/9.8/10.5	2.4/10.5/11.8	2.4/7.9/8.2	1.9/9.3/10.1	2.0/5.3/5.7	3.0/11.2/11.4	10.8/21.7/20.5	8.6/19.2/19.8	1.2/8.7/9.1
lb	2.5/11.6/14.0	2.7/12.5/15.2	2.8/9.3/10.7	2.5/11.1/13.1	2.1/6.4/7.2	2.2/11.0/13.0	3.4/13.5/15.6	2.4/10.2/11.8	1.7/9.0/10.9
oc	4.2/15.2/16.1	4.1/16.5/17.7	6.8/15.3/15.5	6.8/20.3/21.6	3.5/8.3/8.7	5.5/16.4/17.0	5.6/17.2/17.9	3.8/13.5/14.5	3.0/16.1/17.4
ro	3.7/16.1/17.2	3.8/17.1/18.1	5.3/16.5/17.0	4.5/18.6/19.5	3.3/10.0/10.6	3.1/18.3/18.9	8.0/19.4/20.2	5.0/14.3/15.6	2.6/15.4/16.8
sd	0.2/3.2/4.7	0.3/3.7/5.3	0.3/2.4/3.9	0.2/2.9/4.5	0.2/2.4/3.1	0.2/4.0/5.5	0.5/7.0/8.0	0.8/5.7/8.3	0.1/2.4/4.1
sr	2.4/14.3/15.1	3.5/15.4/16.5	3.8/12.2/12.9	3.1/14.4/15.3	3.2/9.3/9.8	4.7/18.5/18.9	7.7/17.8/17.8	5.6/14.0/15.0	1.8/12.6/13.5
ti	1.3/5.9/6.4	1.7/6.8/7.4	1.6/5.1/5.6	1.4/6.6/7.0	1.4/3.7/4.0	2.0/7.8/7.8	7.1/13.8/13.7	5.1/10.3/10.0	1.2/5.5/5.9
uk	3.2/13.0/13.9	2.8/14.5/15.3	3.3/11.3/12.2	3.2/13.4/14.0	3.1/9.4/9.8	5.8/17.4/17.6	5.8/16.7/16.9	4.1/12.8/13.8	1.8/10.5/11.9

SacreBLEU	ha	af	am	gu	lb	oc	ro	sd	sr	ti	uk
en	14.7/16.1/16.8	36.2/34.7/36.2	7.5/9.0/9.4	26.8/26.8/26.7	13.9/14.5/15.9	24.9/24.8/24.9	24.8/26.9/27.6	4.2/4.2/4.5	23.7/22.7/22.7	4.2/4.9/5.5	21.2/19.8/19.7
de	1.9/8.4/9.9	2.6/19.4/20.8	0.4/3.8/4.5	2.1/14.3/15.6	2.2/12.1/13.2	2.4/13.3/13.8	2.3/16.0/17.6	0.5/2.0/2.6	1.4/12.4/13.8	0.2/2.4/3.5	0.8/11.3/12.0
nl	1.7/7.9/8.7	2.5/15.6/17.3	0.4/2.9/3.7	2.3/13.5/14.5	1.9/7.9/9.3	2.5/11.4/11.7	2.3/12.8/14.4	0.5/1.7/2.4	1.2/9.5/10.4	0.2/2.0/2.9	0.7/8.6/9.6
fr	1.8/9.1/10.4	2.5/18.7/20.0	0.3/3.5/4.3	1.5/14.9/16.5	1.8/9.0/10.6	2.7/17.7/18.3	2.4/18.9/20.8	0.4/1.9/2.9	1.1/12.8/14.2	0.2/2.3/3.4	0.7/11.9/12.5
es	1.6/7.1/8.0	1.9/12.6/13.5	0.3/2.8/3.6	1.8/12.2/13.4	1.4/6.7/7.9	2.2/12.7/12.7	2.2/13.4/14.7	0.4/1.7/2.3	1.1/9.8/10.7	0.2/1.9/2.8	0.6/8.6/9.2
ru	1.3/7.0/8.3	1.9/14.0/14.5	0.3/2.9/3.5	2.1/12.7/13.7	1.3/6.8/7.8	1.8/10.5/11.0	1.7/12.8/14.2	0.5/1.8/2.4	1.2/11.7/12.4	0.2/2.1/2.6	1.0/14.4/14.9
cs	1.7/8.3/9.4	2.5/16.6/17.4	0.4/3.4/4.1	1.5/13.7/15.1	1.8/8.3/9.7	2.2/11.8/12.5	2.2/14.5/16.2	0.5/1.9/2.6	1.4/13.2/13.9	0.2/1.9/3.0	0.7/11.7/12.2
hi	1.7/8.0/8.7	2.0/14.4/15.8	0.4/3.7/4.5	2.9/20.5/21.1	1.5/6.3/7.1	1.7/10.0/10.1	1.7/11.5/12.9	0.5/3.1/3.9	1.1/9.5/10.3	0.2/2.4/3.2	0.6/7.8/8.8
bn	1.5/7.0/7.6	1.9/11.8/12.9	0.4/3.1/3.7	2.8/17.6/18.3	1.4/5.3/6.2	1.6/8.5/9.5	1.7/10.1/11.3	0.4/2.6/3.4	1.0/8.4/8.7	0.2/2.0/2.9	0.7/7.2/8.1
ar	1.2/6.9/7.8	1.6/13.6/14.5	0.3/3.0/3.4	1.6/11.4/12.2	1.1/5.6/7.0	1.6/9.7/9.9	1.5/11.8/13.0	0.4/1.9/2.6	1.1/9.8/10.6	0.2/1.9/2.6	0.6/8.3/9.2
he	1.3/7.9/9.0	1.7/16.0/17.3	0.2/3.1/3.8	1.7/12.6/13.5	1.0/6.4/7.8	1.5/11.2/11.7	1.6/13.4/14.7	0.4/1.9/2.7	1.1/11.2/12.0	0.2/1.8/3.0	0.7/9.7/10.1
sv	1.8/9.5/10.4	2.5/20.1/21.7	0.3/3.8/4.4	2.1/14.9/16.0	2.0/9.9/11.2	2.3/13.7/14.5	2.3/16.7/18.4	0.4/2.2/2.8	1.0/12.6/13.6	0.2/2.2/3.3	0.7/11.6/11.7
da	1.9/9.4/10.3	2.6/20.8/22.1	0.4/3.9/4.6	2.0/14.8/16.1	1.8/10.2/11.4	2.3/14.3/14.7	2.3/16.6/18.3	0.4/2.3/3.0	1.2/13.2/14.5	0.3/2.3/3.3	0.8/11.7/12.4
it	1.7/7.7/8.6	2.3/14.1/15.4	0.3/3.0/3.4	2.0/13.1/14.5	1.7/7.1/8.3	2.4/13.9/13.6	2.3/15.0/16.2	0.4/1.7/2.3	1.1/10.5/11.2	0.2/2.0/2.9	0.7/9.5/9.7
pt	1.9/9.4/10.5	2.4/18.6/19.9	0.3/3.8/4.5	1.8/14.6/16.3	1.6/8.5/10.0	2.6/18.1/17.8	2.3/18.4/20.4	0.4/2.1/2.7	1.1/13.0/14.2	0.2/2.3/3.2	0.8/11.7/11.8
pl	1.5/6.7/7.9	2.2/12.1/13.5	0.3/2.9/3.5	2.2/12.3/13.0	1.6/6.6/7.7	2.3/10.0/10.1	2.1/12.5/14.0	0.4/1.6/2.1	1.4/9.5/10.5	0.2/1.8/2.5	0.8/10.1/10.4
bg	1.6/8.1/9.1	2.1/17.2/18.1	0.4/3.5/4.1	2.1/14.4/15.4	1.6/7.9/8.6	2.1/12.6/13.0	1.9/15.5/17.2	0.5/2.1/2.9	1.6/13.8/14.6	0.3/2.2/3.1	0.9/12.5/13.2
kn	1.5/6.4/6.9	1.8/10.2/11.5	0.5/3.1/3.9	4.0/16.2/16.8	1.3/4.8/5.3	1.5/7.8/8.2	1.5/8.9/10.6	0.5/2.6/3.2	1.1/7.0/7.7	0.2/1.9/2.5	0.7/5.6/6.3
mr	1.5/6.5/7.4	1.9/12.0/13.0	0.4/2.9/3.4	4.1/18.9/19.7	1.4/5.2/6.1	1.8/8.2/8.7	1.8/9.3/10.9	0.4/2.2/3.0	1.2/7.6/8.6	0.2/2.0/3.1	0.7/7.0/7.3
mt	2.4/11.4/12.3	3.1/20.7/22.2	0.4/4.3/4.7	2.2/15.4/17.2	2.0/9.0/10.5	2.9/17.3/17.3	2.4/18.8/20.4	0.6/2.5/3.3	1.4/13.9/14.9	0.3/2.4/3.7	0.8/11.4/11.9
ha	-	1.1/10.8/11.4	0.1/2.2/3.0	0.6/8.4/9.9	0.7/5.3/5.9	0.6/7.7/7.9	1.3/8.0/9.3	0.1/1.3/2.3	1.0/6.5/7.6	0.1/1.5/2.3	0.1/6.0/6.4
af	2.5/9.3/10.2	-	0.5/3.7/4.6	3.0/14.4/15.9	2.1/9.4/11.1	3.1/14.1/14.7	2.5/15.0/16.5	0.6/1.9/2.8	2.7/11.6/13.0	0.3/2.3/3.2	1.0/10.2/10.6
am	1.6/4.9/5.9	2.4/9.0/9.7	-	3.0/9.9/11.0	1.4/3.5/4.4	1.9/5.9/6.3	1.9/6.7/7.6	0.6/1.4/2.0	1.8/5.7/6.4	0.8/3.0/3.8	1.0/5.1/5.6
gu	1.4/5.9/6.6	1.7/10.5/12.0	0.4/2.6/3.3	-	1.3/4.7/5.6	1.6/6.9/7.7	1.5/8.3/9.6	0.5/2.4/3.2	1.6/6.8/7.3	0.3/1.6/2.4	1.1/5.8/6.6
lb	1.8/5.8/7.0	2.4/12.9/15.7	0.3/2.1/2.6	1.6/9.4/11.4	-	2.0/8.8/10.1	1.9/8.8/11.7	0.4/1.3/1.7	1.7/6.9/8.5	0.2/1.2/2.0	0.7/6.2/7.4
oc	2.3/8.1/9.4	3.5/16.3/17.6	0.4/3.1/3.6	2.8/12.6/14.0	2.1/7.8/9.3	-	3.5/15.5/17.2	0.5/1.8/2.5	2.8/10.3/11.2	0.3/1.6/2.9	1.0/8.8/9.6
ro	1.7/8.4/9.3	2.7/15.9/17.5	0.3/3.4/3.7	3.0/13.6/15.1	1.8/7.9/9.3	3.0/14.4/14.4	-	0.4/2.2/3.1	1.8/12.0/13.0	0.3/2.0/2.8	0.8/9.9/10.9
sd	0.5/2.3/3.8	0.2/4.1/6.3	0.0/0.9/1.8	0.3/5.3/8.3	0.1/2.1/3.3	0.1/2.5/3.9	0.2/2.6/4.6	-	0.5/2.7/4.0	0.1/0.6/1.4	0.1/2.7/3.6
sr	1.4/7.4/7.9	2.3/14.6/15.7	0.4/3.1/3.7	2.9/12.1/13.4	1.3/6.4/7.8	2.4/10.9/11.1	1.9/13.3/15.0	0.4/1.6/2.5	-	0.3/1.9/2.7	0.9/10.7/11.1
ti	1.1/4.5/4.6	1.8/7.5/8.1	0.5/4.0/3.7	2.1/8.9/9.5	1.0/3.1/3.5	1.1/5.2/5.2	1.4/5.5/6.3	0.5/1.5/1.7	1.2/4.8/5.8	-	0.7/4.2/4.2
uk	1.2/6.3/7.5	2.2/13.2/14.1	0.3/2.5/3.1	2.5/11.5/12.7	1.3/6.1/7.0	2.3/9.7/10.1	2.1/11.7/13.3	0.4/1.6/2.3	1.8/10.8/12.1	0.2/1.5/2.4	-

Table 7: The SacreBLEU performance on EC30 for the baseline, boost-100, and boost-all models in the two-tag fashion, respectively. 970 directions of results are shown, including both English-centric and zero-shot ones. Overall, in all 870 zero-shot directions, the boost-100 model achieves better performance compared to the baseline. Moreover, in 702 out of 870 zero-shot directions, the gain exceeds 5.0.

COMET	en	de	nl	fr	es	ru	cs	hi	bn	ar	he	sv	da	it	pt	pl	bg	kn	mr	mt	ha	af	am	gu	lb	oc	ro	sd	sr	ti	uk
en	-	82.8/82.1/82.2	84.2/83.7/84.0	82.8/81.5/82.0	82.0/81.6/81.5	81.9/81.4/81.3	84.5/83.9/84.0	78.1/78.4/78.3	84.5/84.9/84.8	79.0/79.0/79.1	83.3/83.8/83.7	85.4/84.2/84.2	85.3/84.4/84.4	84.0/83.2/83.4	85.1/84.5/84.4	82.0/81.5/81.6	87.2/86.9/86.6	87.5/87.5/86.0	74.4/75.2/75.5	70.8/71.0/70.8	68.7/72.1/73.4	84.0/83.3/83.0	78.2/80.9/81.0	85.1/85.4/85.3	53.0/53.7/54.0	66.0/65.0/66.0	80.2/78.5/79.3	47.5/48.9/51.8	82.1/80.5/80.7	80.5/81.1/81.9	80.7/79.7/80.0
de	86.4/85.9/85.6	-	61.0/78.6/78.1	62.5/70.9/71.1	64.7/71.3/71.8	45.8/72.5/73.1	56.1/75.6/77.4	36.4/68.1/68.2	36.4/77.6/78.1	46.5/68.1/69.7	53.7/71.4/72.5	66.8/76.5/76.8	66.0/76.5/76.8	61.8/73.3/74.1	65.7/73.5/74.5	55.2/73.2/74.4	49.5/76.6/76.8	42.2/82.1/81.6	37.5/66.6/67.2	36.1/63.1/63.8	47.8/65.0/66.9	62.7/75.3/75.3	41.7/66.8/69.1	41.5/75.9/76.9	41.4/53.0/53.5	55.3/58.1/58.6	59.4/68.1/70.1	39.4/39.6/45.2	55.5/69.5/71.3	32.3/73.4/75.6	46.3/67.5/69.4
nl	84.7/84.5/84.1	55.2/74.7/74.8	-	61.0/70.6/70.7	63.5/71.3/71.3	44.6/71.6/72.4	54.6/74.5/75.1	35.6/68.0/68.2	36.2/77.8/78.3	45.7/66.9/68.7	52.4/71.1/72.2	65.0/74.7/74.8	64.6/75.2/75.3	60.7/72.2/73.1	64.2/72.6/73.1	54.8/72.6/73.2	48.2/75.8/76.1	43.3/82.4/81.9	37.9/66.4/67.1	35.5/62.1/62.7	46.9/64.6/66.8	60.9/78.0/77.2	40.6/66.7/69.0	41.4/76.3/77.1	40.1/50.6/51.2	54.2/57.0/57.3	57.9/67.0/69.0	38.4/39.8/44.5	54.1/68.2/69.3	32.3/72.9/75.5	45.3/66.9/68.4
fr	85.8/85.3/85.1	55.4/71.3/72.0	59.8/75.4/75.9	-	63.8/76.3/76.7	44.4/73.4/74.0	55.0/74.7/75.6	32.6/68.8/68.7	33.7/77.5/78.1	45.5/70.8/71.9	52.1/73.4/74.4	65.7/73.5/74.4	64.6/74.0/74.4	61.1/78.4/78.8	64.9/78.3/78.7	54.6/73.0/74.0	48.2/77.2/77.7	39.2/82.5/82.0	35.4/66.6/67.3	35.3/63.9/64.3	46.8/65.5/67.4	61.8/72.6/72.7	40.5/67.7/70.3	39.6/76.8/78.2	39.7/50.1/50.6	55.2/63.0/63.4	58.5/72.4/73.9	38.4/39.3/45.3	54.6/69.8/71.1	30.5/72.8/75.7	45.0/68.8/70.7
es	83.3/82.8/82.7	53.8/68.7/69.4	58.2/72.7/73.6	60.1/73.7/74.0	-	43.5/71.8/72.8	53.2/72.3/73.8	33.6/68.5/69.0	34.6/78.3/78.8	45.1/69.5/70.6	51.2/70.8/72.0	63.6/71.7/72.7	62.8/72.3/72.9	60.2/77.2/77.4	63.3/80.3/79.9	53.4/71.7/72.7	47.1/75.3/76.2	42.1/82.6/82.3	37.3/66.8/67.5	35.5/62.9/63.2	46.2/64.1/67.0	60.2/69.4/70.0	40.1/67.4/69.8	40.2/76.8/77.8	39.5/49.9/50.3	54.9/62.0/62.4	57.1/71.3/72.4	37.8/39.7/45.6	53.4/67.9/69.5	31.2/73.5/75.8	44.0/67.2/69.0
ru	82.2/81.8/81.3	53.6/68.4/68.8	58.0/71.1/72.0	59.0/68.5/68.4	61.4/69.4/70.0	-	53.4/74.9/75.3	34.2/66.2/66.3	36.6/76.3/76.6	45.1/67.1/68.3	51.6/69.2/70.2	62.8/70.6/71.4	62.1/70.9/71.4	59.0/70.8/71.6	62.3/71.3/71.5	53.5/73.1/73.8	50.8/82.2/82.1	45.1/81.5/80.5	38.8/65.2/65.7	36.8/62.2/62.6	46.7/63.2/65.2	59.7/67.9/68.6	41.3/66.0/67.9	42.5/74.6/75.5	40.2/50.3/50.7	53.5/57.4/58.0	56.6/65.7/67.7	38.8/39.1/43.7	53.8/69.9/70.6	33.6/72.9/75.6	46.5/78.3/78.7
cs	85.0/84.7/84.3	55.3/72.3/73.3	59.5/74.9/75.7	60.8/70.4/70.9	63.4/71.6/72.3	45.2/75.0/75.3	-	34.1/67.3/67.5	34.5/77.1/77.5	46.2/67.9/69.3	52.6/71.2/72.0	65.2/74.1/74.5	64.4/73.9/74.1	60.3/72.9/73.9	64.1/73.8/74.2	54.9/75.8/76.1	48.9/78.4/78.4	39.6/82.2/81.7	34.8/65.8/66.8	36.7/63.5/63.6	47.6/65.7/67.3	61.2/71.6/72.4	41.3/67.1/68.9	39.7/75.5/76.4	40.5/51.7/52.3	54.7/58.6/59.6	58.2/68.2/70.1	39.0/39.3/44.5	54.8/72.4/73.6	31.5/72.9/75.7	46.0/70.4/71.5
hi	85.1/84.8/84.6	58.3/67.9/68.6	62.6/71.6/72.6	63.4/68.8/69.2	66.0/70.9/70.9	48.3/70.7/71.1	57.3/71.4/72.1	-	43.6/82.3/82.7	50.4/67.2/68.8	56.1/69.5/70.8	67.1/71.7/72.3	66.3/72.5/73.2	63.0/70.2/71.0	66.5/72.9/73.3	57.0/69.1/70.1	51.9/72.8/73.6	48.7/86.4/85.5	44.3/73.6/73.7	41.4/67.3/67.9	51.6/67.3/69.1	64.0/71.0/71.6	45.9/68.2/70.1	48.0/81.8/82.2	45.3/54.5/55.4	58.3/61.5/62.3	61.0/65.8/68.0	43.4/44.6/49.3	57.9/67.3/69.4	35.5/74.7/76.9	49.2/65.1/67.6
bn	82.4/81.9/81.3	55.2/64.5/65.5	59.5/69.0/69.9	60.2/64.8/65.7	63.4/67.8/68.3	46.5/68.0/68.4	54.8/67.7/68.6	41.5/70.7/70.5	-	48.5/64.9/66.8	53.9/67.5/68.5	64.6/68.9/69.8	63.8/69.6/70.1	60.0/67.4/68.2	63.8/69.1/70.2	54.4/65.6/66.9	49.6/70.5/71.5	47.4/85.4/84.8	41.9/70.1/70.5	40.5/65.6/66.4	49.6/64.2/66.8	61.2/68.3/68.8	44.3/66.7/68.6	45.8/79.4/79.7	43.6/53.4/54.8	56.2/59.7/60.7	58.1/63.7/66.0	41.5/42.6/47.1	55.5/64.5/66.4	34.2/74.0/76.8	47.3/63.5/65.6
ar	79.8/78.9/78.8	51.1/61.9/62.7	55.9/65.8/66.8	56.7/64.8/65.1	59.6/66.9/67.2	44.1/66.3/67.3	51.4/65.3/66.6	36.0/64.0/64.0	38.7/75.5/75.7	-	51.0/66.9/68.1	60.2/65.9/66.5	59.5/66.6/67.0	56.3/66.2/67.0	59.5/67.3/68.1	51.0/63.7/64.9	47.3/68.9/69.5	48.4/81.0/80.4	40.1/63.6/64.1	38.5/63.0/63.1	47.0/61.6/64.1	58.1/64.4/65.1	41.7/63.7/66.4	42.8/71.8/73.2	40.8/49.5/50.3	53.4/56.2/57.1	54.4/61.5/63.5	39.3/41.1/45.3	52.1/61.5/63.4	34.8/71.8/74.5	44.3/61.1/63.5
he	83.8/83.0/82.7	55.4/66.9/67.0	59.4/70.1/70.7	60.7/67.6/68.4	63.0/69.0/69.2	45.7/68.6/69.7	54.4/69.0/70.6	36.8/65.4/65.9	38.5/76.2/76.7	47.4/68.1/69.1	-	64.7/69.0/69.7	64.2/69.8/70.2	59.9/68.7/69.8	64.0/70.3/70.7	53.9/67.1/68.1	49.2/72.4/73.0	43.3/81.5/80.7	39.0/64.1/65.1	37.5/63.9/64.1	47.8/63.3/66.2	61.5/68.0/68.8	42.3/66.6/68.8	42.6/73.8/75.1	41.4/50.5/51.2	54.8/57.7/58.7	57.6/64.5/66.4	39.8/39.8/44.4	54.9/65.0/66.2	33.4/74.6/76.8	46.2/63.5/65.9
sv	85.5/84.8/84.6	55.3/72.0/72.9	59.4/74.8/75.2	60.4/69.2/69.9	62.9/70.9/71.5	44.8/71.6/71.7	54.4/73.8/74.6	38.6/68.7/69.3	38.1/77.7/78.5	46.0/67.7/69.2	52.7/70.0/72.0	-	65.1/79.8/79.5	60.2/72.0/72.6	64.2/73.0/73.9	54.0/71.0/72.1	48.4/75.3/75.7	44.7/83.1/82.7	38.6/67.5/68.4	36.2/64.3/64.7	47.5/64.9/67.0	61.6/73.0/73.6	41.1/68.3/69.8	41.6/76.7/77.9	40.6/51.3/52.2	54.8/58.1/59.2	58.0/67.6/69.3	38.9/40.2/45.8	54.7/68.6/70.5	32.0/74.3/76.2	45.4/66.7/68.0
da	86.2/85.5/85.3	55.7/72.9/73.1	60.2/75.5/75.5	60.8/69.7/70.0	63.4/71.0/71.4	45.1/71.7/72.0	54.7/73.9/74.7	37.3/69.3/69.4	37.6/78.8/79.1	46.6/67.7/69.6	53.0/70.7/72.2	66.2/80.0/79.6	-	60.6/71.8/72.7	64.5/73.5/73.8	54.6/71.1/71.6	48.6/76.1/76.4	44.4/83.4/82.6	39.0/67.9/68.7	36.1/64.3/64.8	47.7/66.4/67.8	62.2/74.5/74.8	41.1/68.6/70.1	41.5/77.3/78.4	40.5/51.5/51.9	54.9/58.3/59.0	58.1/67.9/70.3	38.9/41.3/46.2	54.8/68.8/70.3	31.5/74.1/76.4	45.3/67.2/68.6
it	84.4/83.9/83.7	54.0/69.6/70.6	58.8/73.6/74.3	60.8/74.7/75.0	63.2/76.4/76.7	44.6/72.4/72.8	54.1/74.0/74.8	36.5/67.7/67.5	37.4/77.4/77.8	45.5/68.8/70.5	51.9/71.6/72.5	64.0/73.4/74.0	62.9/73.4/73.9	-	63.6/78.1/77.7	54.0/72.7/73.3	48.0/76.6/76.2	44.1/82.1/81.3	38.3/66.0/66.8	36.4/63.2/63.4	46.7/64.7/67.2	60.5/70.6/71.2	40.4/67.5/69.4	41.5/76.0/77.0	40.1/49.9/50.4	54.6/61.5/61.6	57.6/72.0/73.4	38.4/39.8/44.7	53.7/69.4/70.4	31.4/73.3/75.9	44.8/67.7/68.8
pt	85.5/85.1/84.8	54.6/70.2/70.4	59.1/73.6/74.4	61.1/74.6/74.5	63.6/78.7/78.6	44.0/71.8/72.1	53.7/72.8/74.3	36.5/69.7/70.3	36.5/79.0/79.2	45.5/68.9/70.5	52.0/71.4/72.7	64.6/73.1/73.6	64.3/73.9/74.2	60.4/77.3/77.8	-	53.8/71.4/72.1	47.8/75.8/76.8	41.5/83.4/83.0	37.0/67.8/69.1	35.4/65.2/65.6	46.6/65.6/67.5	61.3/71.7/72.1	40.3/69.4/71.0	40.6/77.9/78.8	39.9/50.8/51.2	55.3/62.7/63.5	57.7/71.1/72.4	38.2/41.3/46.2	54.2/68.7/70.3	30.6/74.1/76.5	44.4/67.1/68.7
pl	81.2/80.9/80.5	51.5/69.2/69.9	56.2/72.1/72.8	57.3/68.7/68.7	59.9/70.3/70.4	43.2/73.5/74.3	51.7/75.3/76.2	35.3/66.0/66.0	37.0/75.7/75.8	44.2/66.2/67.8	49.8/68.3/69.4	60.9/71.2/71.5	59.8/71.1/71.1	57.1/71.5/72.7	60.2/71.6/72.2	-	46.9/75.8/76.3	45.4/80.9/80.4	38.2/64.9/65.1	35.9/61.3/61.8	45.5/64.2/66.0	57.6/68.1/68.2	39.4/64.5/66.8	41.5/74.0/74.7	39.1/49.0/49.4	52.2/57.0/57.5	55.0/66.8/69.0	37.1/39.2/43.8	51.7/69.8/71.1	32.8/71.9/73.8	43.4/69.5/70.4
bg	85.3/84.9/84.6	56.2/71.0/71.4	60.6/73.9/74.6	61.8/71.5/71.6	64.2/72.4/72.8	48.3/80.1/79.8	55.5/75.9/76.6	38.1/67.0/67.3	38.9/77.5/77.8	48.0/68.2/69.8	53.7/71.6/72.7	66.0/73.7/74.2	65.3/73.7/73.9	61.4/73.8/74.3	65.3/74.2/74.8	55.7/73.8/73.8	-	46.2/82.1/81.4	40.0/65.9/66.5	37.5/64.2/64.6	48.1/65.3/68.3	62.4/71.9/72.0	42.5/68.7/69.9	43.3/75.5/76.6	41.5/51.5/51.9	55.7/59.7/60.4	59.0/69.2/70.6	40.2/40.8/44.3	56.2/72.6/73.9	34.4/74.3/76.4	47.8/75.8/76.6
kn	79.6/79.3/79.1	52.8/62.2/62.9	57.6/67.3/68.2	57.9/62.8/63.2	60.5/65.9/66.3	45.6/66.0/66.9	52.9/65.5/66.6	45.2/69.1/68.7	46.5/80.3/80.3	47.4/64.4/65.6	52.2/65.2/66.5	61.6/67.2/68.2	61.0/68.0/68.3	57.6/65.3/65.9	60.8/67.5/67.9	51.9/64.1/64.9	48.6/68.4/69.1	-	47.1/69.2/69.2	40.5/66.4/67.0	48.6/65.0/65.6	59.5/66.7/67.1	43.7/66.9/68.5	49.9/78.9/79.4	42.7/53.1/54.2	54.6/58.3/59.4	56.0/61.3/63.4	41.2/43.8/47.5	53.4/63.3/64.4	37.4/74.0/76.5	46.3/61.5/63.4
mr	81.4/80.9/80.8	54.0/64.1/64.5	58.7/67.8/68.4	58.6/64.3/65.0	61.7/67.0/67.8	46.4/67.3/68.4	54.1/66.9/68.6	48.0/72.6/72.7	46.6/81.4/81.5	47.6/64.6/66.0	53.3/66.0/67.0	62.7/68.5/69.3	61.6/69.0/69.6	58.7/66.3/67.5	62.0/68.8/69.9	52.8/65.7/67.0	49.6/69.0/70.2	56.8/85.8/85.2	-	40.9/65.9/66.8	47.9/62.7/66.2	59.9/67.5/68.5	44.1/65.6/67.8	49.9/80.4/81.2	42.8/53.5/54.9	54.8/59.3/60.5	57.3/63.2/64.8	41.2/42.3/47.3	54.5/64.1/65.8	37.8/73.1/75.6	47.0/62.7/64.9
mt	80.0/79.5/79.4	48.9/63.6/64.6	52.7/66.8/67.8	54.3/65.1/65.7	56.3/66.4/67.0	39.7/66.7/66.7	47.7/66.9/68.4	34.0/66.7/67.1	34.2/77.7/78.0	42.1/64.4/65.7	46.6/66.3/67.6	58.0/67.0/68.3	58.1/68.0/68.9	53.9/67.9/68.4	57.3/69.5/70.0	48.1/65.2/66.3	43.4/70.7/71.1	42.4/83.5/82.8	37.1/66.6/67.0	-	42.3/65.9/66.9	54.5/66.9/67.7	37.3/65.4/67.5	39.1/75.6/76.6	36.8/52.0/52.6	49.2/58.7/59.2	51.1/65.1/66.6	35.7/40.8/45.0	48.3/63.9/65.5	32.2/74.0/76.2	40.3/63.2/64.6
ha	50.9/67.8/68.1	30.0/50.2/52.4	32.8/55.0/57.0	34.5/51.7/53.1	35.5/54.8/56.3	27.0/55.1/57.2	30.7/53.5/55.7	24.7/59.6/60.1	29.0/72.5/73.3	28.2/54.9/58.2	29.9/54.5/57.0	36.1/55.4/56.8	35.5/55.6/57.5	35.2/54.0/56.1	35.6/56.0/57.7	32.4/54.5/56.4	28.6/58.1/60.0	46.4/78.8/79.0	30.5/60.1/61.6	34.2/61.6/62.2	-	35.0/55.9/57.4	26.7/57.6/61.8	34.2/68.6/71.1	29.8/46.2/47.0	32.1/48.8/50.5	33.7/51.9/54.4	24.8/33.8/40.5	33.3/53.4/55.4	30.7/69.4/72.5	27.0/52.3/54.5
af	83.8/83.1/83.3	52.0/66.2/67.8	56.2/74.4/74.6	57.1/65.4/66.4	59.5/65.9/66.1	42.8/66.7/67.7	50.8/68.1/70.0	36.6/68.2/68.4	40.2/78.8/78.8	44.6/64.0/65.9	49.3/66.4/68.6	60.7/69.4/70.0	60.5/70.1/71.1	56.8/66.9/68.3	60.2/68.7/69.8	50.9/65.8/67.8	47.7/70.9/71.4	50.3/83.4/82.7	42.9/67.1/68.3	35.7/64.5/65.0	45.1/64.8/67.2	-	39.0/66.4/69.2	44.9/76.5/77.8	38.6/50.9/51.7	50.9/55.3/56.7	53.3/62.4/64.9	36.2/39.3/45.5	50.1/63.3/65.7	39.2/74.5/76.7	42.4/62.4/64.7
am	77.7/76.8/77.0	49.6/57.8/58.5	54.9/62.0/62.9	53.5/58.5/59.6	57.1/62.4/63.0	47.1/61.7/62.8	51.0/60.6/62.4	42.1/62.6/62.8	52.1/74.9/75.4	45.5/60.2/62.3	51.4/62.9/64.5	56.3/62.4/63.7	56.7/62.3/63.5	55.4/61.6/62.4	57.2/64.0/64.9	50.2/59.7/61.7	50.6/64.5/65.5	64.7/81.1/79.8	48.1/62.8/63.6	44.9/63.3/64.0	47.6/59.1/63.5	56.8/62.4/63.2	-	51.6/72.1/73.0	43.4/50.6/52.0	51.8/55.2/56.5	52.7/57.9/60.1	37.8/38.0/43.3	50.8/58.0/59.8	60.9/79.4/81.2	45.3/58.2/59.8
gu	81.3/80.8/80.8	54.4/62.3/63.1	59.4/66.8/67.3	58.6/63.1/63.8	61.5/66.2/66.8	50.0/65.7/66.5	54.7/64.9/65.9	55.4/71.8/71.8	57.5/81.4/82.1	49.0/63.2/64.2	54.9/64.6/66.6	61.8/67.4/68.1	61.3/67.7/68.3	58.5/64.6/65.8	61.0/67.5/68.7	53.7/63.8/65.2	52.8/68.2/68.8	69.7/86.3/85.5	56.3/71.7/72.0	45.1/67.5/68.0	51.4/64.4/66.3	60.7/66.7/67.5	47.0/65.6/67.7	-	45.7/55.1/56.0	56.6/59.3/61.0	56.7/61.3/63.4	43.1/44.4/48.7	54.2/62.4/64.5	46.9/73.9/76.7	48.7/61.1/63.0
lb	60.8/62.4/65.1	39.0/52.9/56.3	41.0/54.7/58.0	41.7/50.3/52.8	43.5/53.1/55.3	33.5/55.0/57.3	36.4/53.7/57.1	33.0/61.5/62.7	36.5/74.5/75.9	33.8/52.6/56.2	35.8/54.0/57.4	43.7/54.9/57.9	43.7/55.6/58.0	42.2/52.6/55.8	43.3/53.8/56.9	38.4/52.7/55.4	36.4/57.1/60.1	47.8/79.5/80.5	39.5/62.7/64.4	34.8/63.2/64.3	38.2/57.4/59.5	42.3/54.8/57.9	31.6/57.8/61.2	40.0/71.4/73.6	-	41.3/50.2/52.2	38.3/50.6/54.5	29.7/37.7/41.8	38.5/51.4/54.5	36.4/70.1/73.3	32.4/51.7/55.2
oc	74.4/74.4/74.7	44.7/58.7/59.7	48.9/63.1/63.5	53.9/64.2/64.9	54.9/66.5/66.8	41.3/63.2/64.0	44.2/63.4/65.1	37.2/66.0/66.4	43.0/77.7/78.6	42.1/62.4/63.6	44.3/63.5/65.2	52.1/63.3/64.0	51.6/63.6/64.3	53.6/66.2/67.2	54.8/68.0/68.7	45.9/61.5/62.5	47.5/67.3/67.8	58.2/82.4/82.2	45.6/66.2/66.8	39.0/65.3/65.9	41.5/60.4/63.0	49.2/61.6/62.8	36.1/64.3/66.6	46.2/75.7/76.9	36.1/50.2/50.8	-	47.0/62.1/64.0	33.4/39.2/44.2	44.5/60.1/61.7	42.6/73.4/76.3	38.1/59.2/60.8
ro	81.7/81.0/81.3	51.7/66.1/67.0	56.3/70.0/70.6	58.8/69.6/69.7	60.5/70.8/71.1	44.7/68.5/69.2	51.8/69.2/70.9	41.8/66.2/65.8	43.7/76.6/77.2	46.0/65.5/67.3	50.5/68.4/69.8	59.8/69.9/70.9	59.4/69.9/70.6	58.8/71.9/73.0	60.7/72.5/73.2	52.3/68.7/69.7	48.1/73.1/73.7	56.9/81.2/81.2	45.0/64.5/65.5	39.6/64.3/64.3	46.5/64.8/66.0	57.8/67.9/68.7	40.1/66.4/67.4	44.5/74.0/75.5	40.1/50.0/50.3	54.1/59.2/60.1	-	37.6/39.7/43.9	51.6/66.1/68.1	39.7/73.2/75.1	43.2/63.4/65.6
sd	41.3/55.3/59.4	24.5/40.7/44.7	25.9/45.4/49.8	28.7/42.2/46.2	29.0/46.9/50.1	25.7/46.6/49.6	26.2/43.9/48.3	24.3/55.2/58.8	29.7/68.6/71.8	25.9/46.1/51.1	25.2/45.2/50.2	29.6/46.9/50.3	30.5/47.1/50.6	29.6/45.5/48.9	29.2/47.4/51.1	28.7/45.2/49.0	26.0/47.5/51.0	49.2/75.4/75.3	36.6/57.5/60.6	35.2/55.4/57.8	40.3/49.7/55.2	30.5/47.1/51.3	24.9/48.1/55.6	36.7/64.8/69.6	29.1/45.5/47.6	29.5/45.9/48.9	28.0/42.7/47.4	-	30.7/45.3/48.4	32.1/60.6/69.5	23.2/44.6/47.9
sr	80.8/79.7/79.6	50.9/64.5/65.3	55.4/68.1/69.1	56.0/64.0/64.5	58.7/66.3/67.5	46.1/68.8/70.0	51.0/70.6/71.6	42.4/66.0/66.0	47.0/75.9/76.1	46.2/63.7/65.2	50.0/65.7/67.0	59.0/67.6/68.5	59.0/68.7/69.2	55.9/66.5/67.9	58.7/67.6/68.5	52.0/68.2/69.5	50.4/74.1/74.3	59.2/81.6/80.7	47.9/65.3/65.8	39.7/63.0/63.0	46.0/63.2/63.9	56.4/65.7/66.7	40.6/65.1/66.3	47.0/73.6/74.5	39.9/49.9/50.1	52.6/56.1/56.7	52.7/63.2/65.2	37.6/39.0/42.7	-	42.5/72.6/74.4	43.5/64.4/65.9
ti	71.0/70.5/71.0	46.9/52.9/54.2	51.2/57.8/58.3	51.6/54.4/55.2	54.1/57.5/58.4	42.7/59.4/60.2	47.1/56.9/58.9	43.0/60.6/60.3	47.8/74.1/74.2	44.9/58.0/58.9	48.4/59.5/60.7	54.2/58.7/59.5	54.5/58.9/59.8	51.8/56.5/58.2	54.7/59.2/60.3	48.0/57.0/57.7	45.9/61.1/62.2	58.4/80.4/79.8	47.5/61.3/61.2	43.2/63.7/63.7	46.8/59.5/61.6	53.7/58.2/59.4	43.2/69.9/69.3	46.2/70.1/71.0	42.8/50.6/50.8	51.6/54.0/54.5	49.8/54.6/56.7	39.2/41.5/43.1	48.5/55.5/57.0	-	42.6/56.0/57.4
uk	80.3/80.1/80.0	50.3/64.7/65.5	54.4/68.2/69.1	55.6/64.6/65.3	57.9/66.8/67.4	53.6/77.3/77.5	50.6/70.2/71.7	41.4/64.6/65.0	46.0/75.5/75.5	45.1/64.0/65.9	50.2/66.3/67.6	57.0/67.0/67.8	56.0/67.5/68.3	55.5/67.5/68.5	57.7/68.7/69.2	51.6/69.3/70.3	55.4/77.8/77.9	56.3/80.9/80.0	45.1/64.0/65.0	40.1/62.0/62.6	45.5/61.7/64.2	55.4/65.0/65.8	40.6/63.7/66.2	44.7/72.9/74.1	39.5/49.6/50.3	50.6/55.5/56.6	52.6/62.0/64.6	37.0/37.6/42.9	50.6/65.4/67.3	43.7/72.5/74.8	-

Table 8: The COMET performance on EC30 for the baseline, boost-100, and boost-all models in the two-tag fashion, respectively. 970 directions of results are shown, including English-centric and zero-shot ones. Overall, in 869 out of 870 zero-shot directions, the boost-100 model achieves better performance compared to the baseline. Moreover, in 636 out of 870 zero-shot directions, the gain exceeds 10.0.

ChrF++	en	de	nl	fr	es	ru	cs	hi	bn	ar	he	sv	da	it	pt	pl	bg	kn	mr	mt	ha	af	am	gu	lb	oc	ro	sd	sr	ti	uk
en	-	57.9/57.0/56.6	52.1/51.6/51.8	62.9/61.9/61.4	49.6/49.3/49.1	48.5/47.3/47.1	51.2/49.7/49.2	54.4/54.2/53.4	50.8/50.5/50.6	46.9/46.1/46.1	52.2/51.9/51.3	60.3/58.7/58.2	60.5/59.4/58.7	51.9/51.5/51.5	59.5/58.6/57.9	43.9/43.6/43.4	61.3/60.3/59.6	52.0/51.0/50.6	44.9/45.8/46.5	65.7/65.6/64.7	39.5/42.7/43.7	63.9/62.7/62.0	28.9/30.9/31.2	46.9/47.0/47.6	43.1/44.7/45.8	53.8/53.2/52.9	51.9/52.1/52.5	19.5/22.4/23.3	51.5/49.6/49.7	22.3/22.8/23.7	47.5/46.5/45.9
de	61.9/61.2/60.8	-	41.5/44.2/44.6	41.4/46.7/48.0	35.0/40.8/41.2	26.7/37.9/38.5	25.3/38.9/39.5	31.8/37.7/38.2	31.2/37.8/38.4	26.2/32.6/33.7	19.9/34.8/35.5	42.3/45.7/46.5	41.3/45.9/46.8	38.5/40.8/41.7	40.0/43.0/43.7	30.8/35.2/35.8	42.2/44.8/45.5	36.8/39.1/38.3	28.9/35.2/36.3	28.9/41.2/42.7	16.9/29.9/31.8	35.1/44.6/45.8	2.0/18.1/19.9	29.5/33.2/34.5	26.8/39.9/41.3	30.4/38.8/39.3	32.4/39.2/41.2	3.5/16.2/18.5	32.6/36.5/37.7	11.7/15.5/17.8	29.9/34.3/35.5
nl	54.4/54.2/54.0	40.9/42.9/43.4	-	40.1/43.2/44.0	36.1/38.3/39.4	28.3/34.8/35.1	25.8/34.9/35.3	31.5/35.8/36.0	31.7/36.3/36.8	27.2/30.7/31.5	21.3/32.2/32.8	39.3/41.0/41.8	38.4/41.6/42.3	36.8/38.3/39.2	38.4/40.2/40.7	30.1/33.3/33.6	39.3/40.8/41.4	35.4/37.3/36.5	28.1/33.6/34.9	29.9/38.7/39.7	17.7/29.6/31.4	39.4/43.2/43.9	3.1/17.3/18.9	28.3/31.2/32.5	25.1/33.9/35.2	32.2/36.6/37.3	32.8/36.1/37.8	5.6/16.1/17.8	31.4/33.4/34.7	12.1/14.8/16.8	29.6/31.6/32.6
fr	61.7/61.0/60.6	41.2/44.2/45.0	39.5/42.5/43.3	-	41.5/43.7/44.0	31.8/38.3/39.1	24.7/38.1/38.8	32.4/39.2/39.4	31.8/38.0/38.8	30.6/34.8/35.2	25.0/36.5/37.0	40.0/44.1/44.6	38.0/44.5/45.2	44.4/45.4/45.9	46.2/47.5/47.9	33.6/35.8/36.5	44.3/45.9/46.3	37.8/39.9/38.9	29.0/35.9/37.2	34.8/44.6/46.1	17.6/30.4/32.7	35.8/44.3/45.1	2.0/18.3/20.3	30.3/34.6/35.4	23.0/34.4/36.3	37.6/45.1/45.2	35.5/43.0/44.4	4.7/17.0/19.5	32.3/37.0/38.5	12.2/15.4/17.5	31.8/35.0/36.3
es	53.5/53.4/53.1	36.6/40.0/40.7	36.5/39.7/40.5	43.3/45.5/46.2	-	29.2/35.3/36.1	21.8/34.3/35.0	30.1/35.7/36.0	30.1/36.0/36.2	28.0/31.4/32.3	21.7/32.3/33.0	36.5/40.0/40.6	35.2/40.5/41.1	40.7/42.1/42.6	44.0/44.8/45.1	31.3/33.8/34.7	39.4/41.5/42.1	35.0/37.4/37.3	26.5/33.6/35.0	32.2/40.2/41.2	16.3/29.1/31.2	32.8/40.1/40.7	1.7/16.2/18.4	27.7/31.4/32.6	21.5/31.9/33.4	34.7/40.0/40.2	32.3/38.4/40.2	3.5/15.6/18.2	29.7/34.3/35.4	11.7/14.3/16.4	28.3/32.0/33.0
ru	54.6/53.6/53.2	35.9/40.3/41.1	35.1/39.1/39.3	38.9/42.7/43.6	34.6/38.2/38.7	-	30.0/36.1/36.5	28.6/34.8/34.7	29.8/34.9/35.4	26.2/30.5/31.4	19.6/32.1/32.3	35.5/39.1/39.6	33.6/39.5/40.0	35.8/38.1/38.7	36.8/39.6/40.3	31.4/34.6/34.8	43.8/45.0/45.3	33.8/36.2/35.1	26.0/32.6/33.8	27.3/37.2/38.5	15.1/28.3/30.1	29.0/38.8/39.5	1.7/15.7/17.4	26.7/30.6/31.8	19.0/30.4/31.9	27.2/35.3/36.2	27.4/36.1/37.9	3.9/14.9/17.4	28.6/35.3/36.4	10.4/13.9/15.9	34.4/38.4/39.6
cs	58.7/57.5/57.3	39.2/42.3/43.2	37.8/40.4/41.1	40.1/44.2/45.7	36.4/39.2/40.2	30.9/38.2/38.9	-	30.8/36.5/36.6	29.9/35.9/36.7	27.5/31.2/32.1	20.5/33.2/33.6	38.1/41.3/42.4	37.1/41.8/42.4	37.0/39.3/40.3	38.7/41.4/42.2	34.7/35.7/36.5	42.3/43.9/44.6	34.9/37.2/36.3	27.3/34.1/35.5	29.7/39.5/41.1	16.2/29.3/31.8	32.4/40.9/41.9	1.9/16.7/18.6	27.0/31.4/33.1	20.7/32.0/34.3	28.4/37.0/37.8	29.4/37.8/39.9	3.5/15.3/17.9	31.9/36.5/38.1	10.7/14.2/16.5	30.0/34.4/35.7
hi	58.1/57.1/56.1	21.0/37.8/38.7	23.2/37.9/38.2	25.1/40.5/41.7	21.3/36.1/37.0	6.8/33.0/33.7	18.0/32.1/33.0	-	27.8/40.9/41.2	5.8/29.6/30.9	5.8/30.0/30.8	22.0/37.6/38.4	20.5/38.2/38.8	23.0/35.6/36.6	22.0/38.2/38.8	17.4/30.1/30.8	16.8/38.5/39.5	33.6/41.1/40.8	29.5/39.3/40.4	17.9/36.8/38.3	14.1/28.9/30.5	19.9/39.0/40.2	1.1/16.5/19.3	25.6/37.9/39.3	17.5/29.0/30.6	20.2/33.8/34.4	19.4/34.1/36.0	2.1/17.0/20.5	17.0/31.7/32.6	3.9/14.1/16.8	9.5/29.2/30.7
bn	53.7/52.5/52.0	19.3/35.6/36.5	22.0/35.7/36.6	22.9/38.0/39.1	20.3/33.8/34.8	5.6/31.1/32.3	16.8/30.0/30.8	16.9/38.5/39.6	-	4.0/27.8/28.9	5.6/28.0/29.2	20.5/35.2/36.2	19.5/35.9/36.7	21.1/33.9/34.8	20.3/35.7/36.5	16.2/28.5/29.5	13.4/36.2/37.3	31.1/38.9/38.8	24.0/36.5/37.7	17.2/33.8/35.2	13.8/27.3/29.1	19.3/36.2/37.6	1.1/15.3/18.3	20.2/34.7/36.4	17.1/27.3/29.5	19.2/31.6/32.6	18.9/32.3/33.9	1.9/15.4/19.0	16.0/29.5/30.9	3.0/13.6/16.2	7.4/27.3/28.9
ar	53.9/52.7/52.5	32.6/36.8/37.7	32.0/36.2/36.8	38.0/40.6/41.4	33.7/36.1/36.8	24.2/33.1/33.4	26.1/31.4/31.8	26.5/32.7/32.8	26.0/32.6/33.9	-	20.4/31.1/31.5	32.2/36.4/37.3	30.8/37.2/37.8	33.7/35.7/36.7	35.3/37.7/38.3	26.9/29.7/30.4	36.2/38.9/39.6	32.0/34.6/34.4	23.3/30.9/32.4	27.8/36.3/37.7	15.7/27.0/28.9	27.8/37.1/38.0	1.8/15.0/16.8	23.7/28.5/30.0	18.2/27.7/29.1	26.4/33.9/34.5	27.1/33.7/35.8	4.5/15.2/17.0	26.3/31.1/32.6	9.9/12.9/15.4	22.5/29.4/30.3
he	59.3/58.2/57.2	32.9/39.6/40.1	32.8/38.8/39.3	37.5/42.8/43.6	33.1/37.4/38.3	21.7/34.9/35.8	24.7/34.0/34.7	24.6/33.9/34.5	26.6/33.9/34.9	24.9/31.4/32.2	-	32.0/39.1/40.1	29.9/39.6/40.3	33.6/37.2/37.9	34.8/39.9/40.3	26.1/31.3/32.0	37.5/40.9/41.6	33.0/35.0/34.8	24.5/31.9/33.6	25.9/37.4/39.3	15.6/28.6/30.6	26.6/39.8/40.9	1.6/15.9/17.9	25.0/29.7/31.4	17.9/29.4/31.1	25.4/35.2/36.2	25.3/35.8/37.8	4.0/15.8/17.1	26.8/32.7/34.5	9.6/13.5/16.2	24.0/31.0/32.4
sv	62.8/61.8/61.9	43.4/45.3/46.2	40.6/42.9/43.4	43.9/46.9/48.1	37.2/40.0/40.8	32.7/37.4/37.7	32.1/37.6/38.6	33.5/37.9/37.9	33.2/37.7/38.9	29.0/32.3/33.4	24.8/34.5/35.2	-	49.2/50.3/50.6	38.9/40.5/41.4	41.5/43.3/43.9	32.8/34.5/35.3	42.8/44.2/45.1	36.7/37.8/37.9	29.3/35.2/36.5	33.3/41.8/43.1	17.1/30.1/32.7	39.1/45.4/46.4	3.1/17.6/19.6	30.0/33.1/34.8	24.0/35.0/37.2	33.8/39.0/39.6	36.0/39.6/41.2	4.9/16.9/18.9	33.5/36.2/37.6	11.7/14.8/17.0	30.8/33.9/35.0
da	63.8/62.7/62.6	44.1/45.6/46.7	41.6/43.3/43.9	44.5/47.1/48.0	38.0/40.3/41.1	34.4/38.0/38.2	32.6/38.0/38.9	34.7/38.2/38.8	33.3/37.5/38.7	30.2/32.6/33.8	26.3/34.4/35.5	50.4/50.6/50.8	-	39.2/40.4/41.5	42.2/43.2/44.1	33.5/34.8/35.4	43.2/44.3/45.3	37.2/38.8/38.4	30.0/35.6/36.7	34.9/42.3/43.8	17.9/31.0/33.5	40.1/45.8/47.0	3.3/18.3/20.1	30.4/33.1/34.7	25.4/35.8/37.6	35.7/39.8/40.6	36.7/39.4/41.8	5.0/16.8/19.0	34.0/36.4/38.0	12.2/15.0/17.4	32.2/34.1/35.1
it	55.8/55.4/55.1	38.6/41.8/42.1	38.5/40.9/41.3	45.7/48.0/48.8	40.9/43.2/43.5	28.8/36.4/36.6	24.7/35.6/36.2	29.4/36.7/36.7	29.7/36.3/37.0	26.7/32.2/32.8	22.1/33.2/33.6	37.3/40.5/41.5	36.2/41.2/41.7	-	43.6/44.5/44.7	32.0/34.3/34.8	40.7/42.9/43.2	35.1/37.3/37.3	27.1/33.9/35.1	34.0/41.5/42.7	16.7/28.3/30.5	33.2/40.7/41.6	1.9/16.8/18.9	27.9/31.8/33.3	22.6/32.6/34.0	35.9/40.4/41.1	32.8/39.7/41.2	4.1/16.1/18.4	30.1/34.7/35.9	11.7/14.3/16.7	28.4/32.8/34.0
pt	64.1/63.0/62.8	41.0/43.6/44.0	39.0/41.9/42.5	48.9/51.1/51.7	43.7/45.0/45.3	34.1/38.1/38.5	27.8/37.5/38.2	32.3/38.6/38.3	31.9/37.9/38.8	30.0/33.9/34.2	24.7/35.1/35.9	41.4/44.0/44.4	39.8/44.5/45.0	43.1/44.1/44.6	-	33.1/34.6/35.2	43.7/45.1/46.1	37.0/39.0/38.0	29.7/35.6/36.4	37.2/45.7/46.8	17.0/30.5/33.1	36.8/44.2/45.0	2.0/18.3/19.9	30.1/33.6/34.8	23.4/34.1/35.6	39.1/44.4/44.7	36.2/42.0/43.7	4.1/17.0/19.7	32.9/36.5/38.0	12.3/14.9/17.0	31.6/34.2/35.7
pl	50.8/50.4/50.0	35.2/38.6/39.2	34.9/38.5/39.1	37.9/41.4/42.3	35.2/37.8/38.4	29.1/35.9/36.1	30.2/35.4/36.0	28.2/33.7/33.9	27.4/33.2/33.9	25.3/29.3/30.0	18.2/30.5/31.1	35.3/37.7/38.1	35.2/37.8/38.7	36.1/37.8/38.4	36.2/38.3/38.9	-	39.0/41.0/41.4	32.2/34.8/34.3	25.8/32.1/33.0	27.3/36.5/37.6	15.5/27.9/29.7	29.7/37.7/38.5	1.5/14.9/17.1	25.6/29.5/30.5	20.7/29.8/31.3	29.0/34.9/35.2	28.1/35.4/36.8	3.4/15.3/16.9	29.7/33.3/34.4	10.3/13.5/15.8	28.4/32.4/33.5
bg	60.9/59.8/59.5	39.8/42.9/43.7	38.2/41.3/41.8	43.0/46.5/47.3	37.8/40.0/40.7	38.5/41.9/41.7	31.3/38.2/38.6	31.3/37.1/37.4	31.3/36.9/37.5	28.9/33.0/33.4	21.8/34.8/35.4	40.0/42.3/42.8	38.1/42.7/43.1	39.2/40.2/41.0	41.5/43.1/43.8	33.3/35.7/36.0	-	36.2/37.9/36.8	28.9/34.3/35.7	30.4/41.0/42.2	15.5/30.3/31.5	32.6/42.0/43.0	2.2/17.7/19.3	29.1/32.4/33.6	21.7/32.9/34.5	32.8/38.8/39.6	33.7/39.1/41.3	3.9/16.7/18.2	35.8/38.0/39.4	11.7/14.7/16.9	35.0/36.9/38.1
kn	50.5/49.6/49.2	19.0/34.1/35.4	20.9/34.5/35.6	23.5/36.2/37.2	20.2/33.5/34.1	6.6/29.5/30.3	16.7/28.4/29.3	18.0/36.6/37.2	25.5/36.3/37.6	3.7/26.0/27.5	5.0/26.8/27.3	20.3/33.6/34.6	19.7/34.3/34.9	20.9/32.4/33.2	20.5/34.5/35.2	16.3/27.2/28.2	13.5/34.5/35.5	-	23.0/35.1/36.9	17.1/32.9/34.2	13.6/26.7/28.2	19.0/34.7/36.0	1.0/15.5/18.1	21.5/32.5/34.2	16.8/26.8/28.1	19.2/30.8/31.8	18.5/30.6/32.7	1.7/15.9/18.5	16.2/28.5/29.6	2.8/13.1/15.8	7.1/25.8/27.2
mr	51.7/51.1/50.4	18.3/34.5/35.6	20.0/35.0/35.9	21.0/36.3/37.6	18.8/33.5/34.4	4.2/30.3/31.2	15.5/28.9/30.2	17.1/39.1/40.0	23.9/38.1/39.2	2.8/26.8/28.1	4.3/27.5/28.6	19.5/34.5/35.6	19.4/35.1/36.0	19.7/32.4/33.6	19.2/34.7/35.5	15.1/27.8/29.0	8.8/34.7/36.2	29.1/39.1/39.3	-	16.1/33.0/34.4	13.6/25.3/28.3	18.9/35.4/37.2	1.1/15.1/17.8	20.0/35.7/37.4	16.5/26.6/28.4	18.6/30.5/31.8	18.6/31.2/32.9	1.6/15.1/18.6	15.5/28.5/30.0	2.4/13.0/15.9	4.6/26.7/28.3
mt	70.0/68.9/68.2	34.9/43.7/44.9	34.5/42.4/43.1	42.9/49.3/50.8	37.5/42.2/43.2	20.2/37.7/38.4	22.6/38.1/38.7	26.5/39.2/39.8	25.7/38.0/39.1	24.0/33.9/35.3	12.8/35.6/37.0	35.6/44.0/45.3	34.5/44.5/45.7	40.1/43.1/44.3	42.0/47.4/48.6	27.6/34.9/35.9	41.2/45.0/46.3	35.1/39.6/39.3	26.3/35.6/37.6	-	15.7/32.3/35.0	26.3/45.7/46.9	1.4/18.1/20.7	26.6/34.3/35.6	20.0/34.2/36.5	28.7/42.0/43.0	29.1/41.1/43.6	2.1/15.9/18.5	24.8/36.7/38.6	8.9/14.5/17.7	22.0/33.9/35.5
ha	31.7/42.1/42.1	22.8/30.7/31.7	22.5/31.2/32.2	22.4/32.3/33.6	21.6/30.5/31.4	13.1/26.3/27.3	14.9/25.7/26.9	14.5/27.7/28.7	14.2/27.3/28.9	12.4/23.9/25.7	8.4/24.7/25.5	19.6/30.3/31.5	18.1/31.0/31.8	21.8/29.3/30.5	22.6/31.1/31.9	17.6/26.0/26.4	18.7/30.5/31.9	20.3/30.1/30.5	15.3/27.1/29.2	16.9/29.7/31.1	-	16.1/32.1/33.4	1.3/11.7/14.6	13.0/24.4/26.9	12.8/24.2/26.1	14.2/28.7/29.9	15.1/28.0/30.3	6.2/11.0/15.7	16.5/25.6/27.3	4.4/10.8/13.9	14.5/23.5/25.0
af	68.2/66.8/65.8	38.2/42.7/44.2	39.4/42.3/43.3	39.5/44.2/45.9	33.4/37.9/39.3	28.9/35.2/36.3	27.3/35.2/36.8	30.1/36.4/37.8	30.4/36.2/37.9	24.0/30.0/31.9	19.7/33.0/34.2	40.7/43.5/44.7	39.7/44.2/45.2	35.3/38.0/39.5	39.0/41.8/43.4	28.9/32.2/33.2	38.9/41.2/43.1	34.1/37.3/37.7	25.9/34.1/36.1	30.7/40.8/43.3	19.2/29.6/31.8	-	3.7/17.0/20.0	27.4/32.0/34.2	24.3/34.5/36.7	31.2/38.0/38.9	32.0/36.8/39.4	7.0/16.6/18.9	31.3/33.2/35.5	10.3/13.8/16.8	28.8/31.2/33.0
am	46.0/44.2/43.7	23.6/31.2/32.0	24.0/31.3/32.1	28.0/32.3/33.3	25.5/30.5/31.1	17.5/26.4/27.1	17.6/25.5/26.0	19.9/29.1/29.8	21.5/29.5/30.3	13.5/23.6/24.6	11.3/25.0/25.4	24.5/29.9/30.9	23.7/30.7/31.4	25.7/29.6/30.2	26.5/30.8/31.3	18.9/25.1/25.5	27.1/31.0/32.0	28.0/32.2/31.4	20.8/28.7/30.0	18.0/28.9/30.4	14.1/23.1/25.7	19.9/31.5/32.4	-	19.0/25.9/27.7	16.0/23.9/24.6	20.1/27.1/28.0	18.9/27.7/29.0	5.4/12.7/14.9	18.0/25.3/26.8	9.8/15.1/17.8	14.2/23.3/24.9
gu	50.7/49.8/49.8	19.5/33.2/34.3	21.2/33.2/34.3	24.9/34.7/36.0	21.6/32.0/32.7	10.4/28.3/29.4	16.4/27.4/28.5	24.7/38.8/39.1	27.8/36.8/37.7	6.7/24.7/26.7	7.0/26.1/27.1	21.3/32.7/33.7	20.3/33.0/34.0	21.7/30.7/32.1	21.3/32.8/33.8	16.7/26.1/27.2	19.8/32.8/34.4	32.6/37.9/37.9	27.6/36.2/38.0	16.2/31.4/32.7	13.4/24.8/27.3	18.8/34.0/35.2	1.1/14.2/16.4	-	16.4/25.9/27.5	19.4/29.2/30.2	18.8/29.2/31.6	2.4/15.9/19.4	17.1/26.8/28.4	3.9/11.8/14.9	7.9/25.2/26.8
lb	47.2/47.9/49.8	36.6/41.3/43.7	31.2/36.1/38.3	28.6/35.7/39.1	24.6/33.0/35.2	15.9/28.6/31.1	18.4/28.6/31.3	18.2/29.0/31.1	19.3/29.5/32.0	10.4/23.1/26.2	9.1/25.7/28.3	26.7/35.3/38.4	26.6/35.7/39.1	26.2/32.4/34.8	25.9/34.1/36.9	21.0/27.4/29.2	27.8/33.5/36.7	24.4/30.8/32.0	17.6/27.4/30.6	21.1/32.0/35.2	16.4/24.5/28.0	24.4/36.5/40.3	1.6/12.2/15.3	17.8/25.7/29.0	-	22.1/31.4/33.8	20.2/30.4/34.3	4.5/12.7/15.9	21.6/27.9/30.8	5.6/10.3/14.1	18.0/25.1/28.1
oc	60.2/59.7/59.5	33.7/40.0/40.8	33.5/39.2/39.8	47.1/49.3/49.9	39.2/41.9/42.0	26.0/33.8/34.2	20.1/33.6/34.8	26.1/34.6/35.1	26.0/34.1/35.0	21.8/29.8/30.7	18.2/32.0/32.4	33.6/40.0/40.8	33.4/40.7/41.5	38.4/41.3/42.1	43.0/45.8/46.3	27.9/31.6/32.3	38.1/40.5/41.4	31.4/34.9/34.7	20.1/31.7/33.3	29.4/41.5/43.1	18.0/28.5/30.2	30.6/40.7/42.2	2.2/15.4/17.0	24.3/29.9/31.6	21.7/31.8/33.5	-	27.7/38.9/41.0	5.8/16.1/17.9	28.1/32.9/34.4	8.9/12.8/15.3	25.1/30.0/31.3
ro	57.8/57.0/56.9	38.1/41.5/42.5	37.4/40.1/40.9	45.5/47.5/48.7	39.5/41.3/41.7	31.5/36.6/37.0	27.3/36.2/36.6	30.0/35.6/36.2	29.1/35.1/36.4	28.1/31.5/32.6	22.5/33.6/34.1	38.9/41.2/42.0	38.8/41.3/42.4	40.8/41.6/42.7	42.9/44.2/45.0	33.1/34.4/34.9	41.8/43.3/44.0	34.3/36.6/36.0	27.6/33.2/34.5	34.3/41.7/43.4	16.6/29.0/31.6	34.5/40.9/42.2	2.2/16.4/18.0	27.8/31.2/32.8	24.1/32.4/34.0	35.7/40.6/41.1	-	3.7/16.7/18.4	32.1/35.0/36.7	10.7/14.1/16.2	28.6/32.5/34.0
sd	16.2/27.0/31.1	10.7/23.8/26.0	9.7/23.9/25.9	10.5/23.1/25.8	10.6/23.3/25.4	4.8/19.5/21.6	7.4/18.1/20.3	9.3/23.3/26.2	9.5/21.9/26.0	3.2/15.5/19.2	4.8/17.3/19.7	7.2/22.1/24.8	9.4/22.4/25.0	10.6/22.0/24.4	8.3/22.3/24.9	8.4/19.2/21.2	7.8/22.2/25.0	13.3/23.9/24.7	13.2/21.2/26.1	8.0/19.3/22.7	15.1/17.5/21.6	9.9/23.9/27.1	0.6/6.1/11.1	9.4/19.9/25.2	8.8/16.9/20.3	7.0/20.5/23.4	8.0/20.1/23.8	-	8.0/19.9/22.0	1.8/6.5/10.9	6.7/17.9/20.1
sr	56.3/55.2/54.8	34.8/39.3/40.6	34.3/37.8/38.9	38.0/41.6/43.3	34.1/37.2/38.1	33.6/36.3/37.1	29.0/35.4/36.5	27.3/33.4/34.1	26.9/32.9/34.4	25.6/29.7/30.6	20.6/31.0/32.1	36.3/38.8/40.1	35.7/39.3/40.4	34.8/36.8/38.0	36.9/39.0/40.3	30.8/32.7/33.6	41.9/43.4/44.3	31.6/34.1/34.5	25.3/31.7/33.3	29.1/37.4/38.9	15.0/28.4/30.2	29.7/38.8/40.2	2.1/15.2/17.4	25.7/29.2/31.2	21.4/29.7/32.0	28.9/34.6/35.6	28.5/35.5/37.9	3.7/16.0/17.1	-	10.1/13.1/15.5	28.4/32.1/34.2
ti	39.5/38.8/38.9	20.5/29.2/30.2	21.3/29.6/30.4	25.2/29.7/30.5	23.4/28.6/29.4	13.2/25.1/25.7	17.0/23.8/24.4	16.5/26.9/27.8	19.7/27.1/28.7	12.0/22.2/23.3	8.1/22.8/23.8	21.1/28.0/28.9	21.7/29.0/29.2	22.6/27.5/28.3	21.6/28.6/29.3	17.7/23.7/24.1	20.7/29.2/29.8	25.7/29.9/28.7	20.9/27.1/28.2	17.6/26.7/27.8	12.9/22.9/24.8	18.3/29.6/30.6	2.1/16.3/17.7	16.6/24.3/25.8	15.6/23.1/23.7	18.1/25.9/26.2	17.3/25.8/27.4	1.9/13.7/14.7	15.6/24.0/25.1	-	8.1/22.1/23.4
uk	53.8/52.5/52.6	35.3/38.2/39.3	34.2/37.4/38.0	39.1/41.4/42.2	34.7/36.7/37.1	40.5/40.9/41.1	30.6/34.3/34.8	26.2/33.1/33.4	27.1/32.4/33.3	24.9/28.4/29.5	21.6/30.5/31.0	34.9/37.6/38.3	34.0/38.4/38.7	34.5/36.0/37.0	36.5/38.3/38.6	31.1/32.9/33.6	41.5/42.8/43.1	32.1/34.2/32.6	24.5/31.0/32.7	27.9/35.4/36.6	15.6/26.2/28.7	29.0/37.4/38.6	1.7/14.5/16.3	23.4/28.4/30.1	19.9/28.8/30.7	29.0/33.6/34.6	26.7/33.9/36.1	5.5/14.9/17.0	29.5/33.1/34.6	8.7/12.3/15.0	-

Table 9: The ChrF++ performance on EC30 for the baseline, boost-100, and boost-all models in the one-tag fashion, respectively. 970 directions of results are shown, including English-centric and zero-shot ones. Overall, in all 870 zero-shot directions, the boost-100 model achieves better performance compared to the baseline. Moreover, in 564 out of 870 zero-shot directions, the gain exceeds 5.0.

SacreBLEU	en	de	nl	fr	es	ru	cs	hi	bn	ar	he	sv	da	it	pt	pl	bg	kn	mr	mt	ha	af	am	gu	lb	oc	ro	sd	sr	ti	uk
en	-	31.0/30.1/30.1	23.4/22.7/22.9	40.6/39.5/39.0	23.1/22.9/23.2	22.4/21.1/21.6	25.0/23.5/23.4	38.6/38.5/37.5	32.0/32.1/31.7	19.2/18.7/19.4	24.6/23.7/22.2	35.8/33.5/33.6	35.9/34.7/34.8	24.4/23.7/23.7	34.4/33.2/33.1	16.2/15.9/16.3	35.7/34.1/33.9	33.6/34.1/30.0	26.6/27.4/27.5	37.9/37.7/37.6	14.7/16.1/16.8	36.2/34.7/36.2	7.5/9.0/9.4	26.8/26.8/26.7	13.9/14.5/15.9	24.9/24.8/24.9	24.8/26.9/27.6	4.2/4.2/4.5	23.7/22.7/22.7	4.2/4.9/5.5	21.2/19.8/19.7
de	37.6/36.3/35.3	-	2.5/16.4/16.5	2.9/21.7/23.4	2.2/15.5/16.5	1.2/13.1/13.8	1.9/14.9/15.5	1.4/20.9/21.4	2.0/18.9/19.9	0.8/9.4/10.2	0.8/10.5/10.5	2.4/20.1/20.8	2.9/21.7/22.5	2.5/15.8/16.1	2.4/17.6/18.6	2.1/11.0/11.9	1.7/20.3/20.1	4.3/21.1/21.0	3.8/16.1/17.5	2.0/15.1/16.3	1.9/8.4/9.9	2.6/19.4/20.8	0.4/3.8/4.5	2.1/14.3/15.6	2.2/12.1/13.2	2.4/13.3/13.8	2.3/16.0/17.6	0.5/2.0/2.6	1.4/12.4/13.8	0.2/2.4/3.5	0.8/11.3/12.0
nl	28.2/27.2/27.5	2.6/15.5/16.0	-	3.1/17.8/18.9	2.0/13.8/14.5	0.9/10.0/10.5	1.9/11.6/12.1	1.3/19.1/19.5	2.0/16.8/18.2	0.6/7.2/8.2	0.9/8.5/8.6	2.3/15.2/15.8	2.8/16.0/17.2	2.3/12.9/13.4	2.4/14.7/15.4	2.1/9.6/9.6	1.4/16.0/16.4	5.0/19.4/19.8	3.4/15.2/16.3	1.8/12.3/13.7	1.7/7.9/8.7	2.5/15.6/17.3	0.4/2.9/3.7	2.3/13.5/14.5	1.9/7.9/9.3	2.5/11.4/11.7	2.3/12.8/14.4	0.5/1.7/2.4	1.2/9.5/10.4	0.2/2.0/2.9	0.7/8.6/9.6
fr	36.3/35.1/35.3	2.5/17.6/18.3	2.4/14.4/14.9	-	2.2/17.9/18.3	0.9/13.2/14.3	1.7/13.9/14.2	0.8/21.7/22.2	1.2/18.7/19.8	0.7/10.5/11.2	0.9/11.6/11.2	2.4/18.2/19.2	2.6/19.0/20.2	2.6/19.1/19.3	2.6/21.7/22.5	1.8/11.0/11.3	1.5/19.6/19.9	3.6/21.4/21.3	3.3/16.3/17.8	2.1/17.6/18.8	1.8/9.1/10.4	2.5/18.7/20.0	0.3/3.5/4.3	1.5/14.9/16.5	1.8/9.0/10.6	2.7/17.7/18.3	2.4/18.9/20.8	0.4/1.9/2.9	1.1/12.8/14.2	0.2/2.3/3.4	0.7/11.9/12.5
es	25.1/24.1/24.2	1.9/12.4/12.8	1.9/11.3/11.7	3.2/19.1/19.7	-	0.9/10.2/10.9	1.4/10.3/11.0	1.0/17.8/18.5	1.2/16.2/17.5	0.6/7.6/8.0	0.8/7.8/7.8	1.8/12.6/13.7	2.1/14.4/15.5	2.4/15.0/15.7	2.3/17.0/17.6	1.7/8.8/9.3	1.3/14.9/15.3	3.9/18.7/19.4	3.4/14.0/15.1	1.7/12.3/13.1	1.6/7.1/8.0	1.9/12.6/13.5	0.3/2.8/3.6	1.8/12.2/13.4	1.4/6.7/7.9	2.2/12.7/12.7	2.2/13.4/14.7	0.4/1.7/2.3	1.1/9.8/10.7	0.2/1.9/2.8	0.6/8.6/9.2
ru	28.1/26.8/26.5	1.8/14.1/14.5	1.5/11.2/11.5	2.5/18.2/18.8	1.8/13.7/14.4	-	1.3/12.2/12.7	1.0/17.4/18.1	1.4/15.9/17.1	0.7/7.9/8.3	0.8/8.4/8.5	1.5/13.6/14.6	1.8/15.1/16.2	1.9/13.2/13.5	1.8/14.0/14.7	1.5/10.2/10.4	2.2/19.1/19.0	4.0/17.6/17.6	3.1/14.0/15.3	1.3/11.2/12.3	1.3/7.0/8.3	1.9/14.0/14.5	0.3/2.9/3.5	2.1/12.7/13.7	1.3/6.8/7.8	1.8/10.5/11.0	1.7/12.8/14.2	0.5/1.8/2.4	1.2/11.7/12.4	0.2/2.1/2.6	1.0/14.4/14.9
cs	33.0/31.6/31.3	2.6/16.2/16.8	2.2/13.2/13.3	3.0/20.1/21.2	2.1/14.0/15.2	1.1/13.9/14.1	-	1.0/20.3/19.8	1.2/17.4/18.6	0.8/8.1/9.1	0.9/9.5/9.3	2.2/16.6/17.5	2.6/17.7/18.9	2.4/13.8/14.2	2.3/16.3/17.0	2.4/11.9/12.2	1.8/19.1/18.8	2.9/20.5/20.6	2.2/15.4/16.5	1.9/13.4/14.6	1.7/8.3/9.4	2.5/16.6/17.4	0.4/3.4/4.1	1.5/13.7/15.1	1.8/8.3/9.7	2.2/11.8/12.5	2.2/14.5/16.2	0.5/1.9/2.6	1.4/13.2/13.9	0.2/1.9/3.0	0.7/11.7/12.2
hi	33.6/31.6/31.0	1.8/12.2/13.0	1.9/10.7/10.9	2.0/16.0/17.0	1.6/11.6/12.3	0.8/9.6/10.5	1.4/10.3/10.6	-	3.1/22.6/23.2	0.6/7.7/8.6	0.7/7.5/7.9	1.8/12.6/13.8	2.0/13.8/15.0	1.8/11.1/11.6	1.8/13.3/13.9	1.5/7.1/7.6	1.2/14.9/15.4	5.6/24.8/24.1	5.4/21.3/22.0	1.7/11.4/13.2	1.7/8.0/8.7	2.0/14.4/15.8	0.4/3.7/4.5	2.9/20.5/21.1	1.5/6.3/7.1	1.7/10.0/10.1	1.7/11.5/12.9	0.5/3.1/3.9	1.1/9.5/10.3	0.2/2.4/3.2	0.6/7.8/8.8
bn	28.1/26.3/25.8	1.9/10.5/10.9	1.9/9.7/10.3	2.0/14.0/14.9	1.5/10.6/11.6	0.8/8.6/8.9	1.3/8.2/8.8	2.3/22.6/22.5	-	0.5/6.4/6.9	0.7/6.9/7.2	2.0/11.1/12.1	2.1/12.6/13.6	1.6/9.7/10.3	1.7/11.3/12.1	1.4/6.8/7.2	1.1/12.8/13.0	5.1/22.7/22.4	4.2/18.9/19.3	1.5/9.3/10.3	1.5/7.0/7.6	1.9/11.8/12.9	0.4/3.1/3.7	2.8/17.6/18.3	1.4/5.3/6.2	1.6/8.5/9.5	1.7/10.1/11.3	0.4/2.6/3.4	1.0/8.4/8.7	0.2/2.0/2.9	0.7/7.2/8.1
ar	28.3/26.7/27.0	1.6/11.8/12.2	1.3/9.8/10.6	2.4/16.5/17.6	1.7/11.6/12.4	1.0/9.5/10.3	1.1/9.6/9.8	1.4/16.6/16.8	1.7/14.2/15.2	-	0.7/8.5/8.1	1.4/12.0/13.4	1.5/13.4/14.4	1.6/11.1/11.5	1.6/12.8/13.7	1.3/7.3/7.3	1.4/14.6/15.2	4.5/16.6/16.9	3.3/12.9/13.8	1.1/11.8/12.5	1.2/6.9/7.8	1.6/13.6/14.5	0.3/3.0/3.4	1.6/11.4/12.2	1.1/5.6/7.0	1.6/9.7/9.9	1.5/11.8/13.0	0.4/1.9/2.6	1.1/9.8/10.6	0.2/1.9/2.6	0.6/8.3/9.2
he	35.9/33.7/32.7	1.4/14.2/14.9	1.5/12.0/11.8	2.3/19.6/21.1	1.5/13.5/13.9	0.9/11.4/12.0	1.2/11.4/11.9	1.3/18.2/18.8	1.7/16.3/17.5	0.6/8.6/9.2	-	1.3/15.2/15.9	1.6/16.0/17.1	1.4/12.1/12.9	1.5/15.3/15.7	1.2/8.4/8.9	1.4/17.3/17.2	3.7/18.3/17.2	3.4/14.6/15.1	1.3/13.1/14.5	1.3/7.9/9.0	1.7/16.0/17.3	0.2/3.1/3.8	1.7/12.6/13.5	1.0/6.4/7.8	1.5/11.2/11.7	1.6/13.4/14.7	0.4/1.9/2.7	1.1/11.2/12.0	0.2/1.8/3.0	0.7/9.7/10.1
sv	39.3/38.2/38.1	2.9/19.7/20.4	2.5/15.7/16.4	3.3/23.1/24.0	2.2/15.5/16.0	1.0/13.6/13.8	1.9/14.0/14.7	2.2/20.7/21.7	2.6/18.6/19.5	0.8/9.1/10.4	0.9/10.9/11.4	-	3.4/26.2/26.7	2.4/15.3/15.8	2.6/19.1/19.7	2.0/10.8/11.2	1.4/19.7/20.5	4.8/20.6/21.0	4.2/16.1/17.5	1.9/16.4/17.9	1.8/9.5/10.4	2.5/20.1/21.7	0.3/3.8/4.4	2.1/14.9/16.0	2.0/9.9/11.2	2.3/13.7/14.5	2.3/16.7/18.4	0.4/2.2/2.8	1.0/12.6/13.6	0.2/2.2/3.3	0.7/11.6/11.7
da	39.5/38.3/38.4	2.9/19.8/19.9	2.6/16.4/16.4	3.2/22.9/24.4	2.3/15.7/16.6	1.1/12.9/13.3	2.0/14.7/15.4	1.6/21.6/21.5	2.4/18.6/19.6	0.8/9.3/10.6	0.9/10.8/10.9	3.2/25.3/25.4	-	2.5/15.6/16.2	2.6/18.8/19.8	2.4/10.8/11.2	1.6/19.9/20.3	5.4/21.0/20.9	4.3/16.4/17.5	2.0/15.8/17.4	1.9/9.4/10.3	2.6/20.8/22.1	0.4/3.9/4.6	2.0/14.8/16.1	1.8/10.2/11.4	2.3/14.3/14.7	2.3/16.6/18.3	0.4/2.3/3.0	1.2/13.2/14.5	0.3/2.3/3.3	0.8/11.7/12.4
it	27.4/26.8/27.0	2.6/14.3/14.7	2.2/12.4/13.0	3.7/22.6/23.4	2.4/16.9/17.6	1.0/11.1/11.7	1.8/11.3/12.2	1.4/19.2/19.4	2.3/16.8/18.1	0.7/7.7/8.5	0.8/8.7/8.5	2.3/13.8/14.9	2.5/15.0/16.2	-	2.7/17.5/18.2	2.0/9.6/10.4	1.4/16.8/16.9	5.1/19.3/19.3	3.7/14.9/15.8	2.0/13.9/15.0	1.7/7.7/8.6	2.3/14.1/15.4	0.3/3.0/3.4	2.0/13.1/14.5	1.7/7.1/8.3	2.4/13.9/13.6	2.3/15.0/16.2	0.4/1.7/2.3	1.1/10.5/11.2	0.2/2.0/2.9	0.7/9.5/9.7
pt	39.5/38.3/38.3	2.4/17.8/18.0	2.3/14.2/14.9	3.9/27.3/28.2	2.5/19.2/19.9	1.1/13.1/13.6	1.7/14.1/14.5	1.4/21.9/22.4	2.1/19.0/20.0	0.9/9.8/10.9	0.8/11.3/11.3	2.4/18.2/19.7	2.5/20.2/21.8	2.5/17.7/17.8	-	1.9/10.6/11.0	1.5/20.2/21.0	4.2/21.3/21.5	3.7/16.1/17.6	2.1/18.8/20.2	1.9/9.4/10.5	2.4/18.6/19.9	0.3/3.8/4.5	1.8/14.6/16.3	1.6/8.5/10.0	2.6/18.1/17.8	2.3/18.4/20.4	0.4/2.1/2.7	1.1/13.0/14.2	0.2/2.3/3.2	0.8/11.7/11.8
pl	23.2/22.3/22.8	2.4/12.2/12.9	2.1/11.2/11.6	3.1/16.9/17.9	2.2/13.2/13.6	1.1/11.7/11.7	2.0/12.2/12.7	1.2/17.5/17.3	2.2/15.0/15.7	0.8/6.5/7.5	0.7/7.3/7.5	2.0/12.3/12.9	2.4/13.5/14.6	2.5/12.3/13.1	2.2/13.0/13.5	-	1.9/15.9/16.0	4.5/17.7/17.9	3.4/13.9/14.6	1.8/10.8/11.8	1.5/6.7/7.9	2.2/12.1/13.5	0.3/2.9/3.5	2.2/12.3/13.0	1.6/6.6/7.7	2.3/10.0/10.1	2.1/12.5/14.0	0.4/1.6/2.1	1.4/9.5/10.5	0.2/1.8/2.5	0.8/10.1/10.4
bg	34.9/33.2/33.3	2.0/16.1/17.1	1.9/13.1/13.7	2.5/21.7/22.5	1.8/14.9/15.4	1.5/15.9/16.0	1.5/13.8/14.2	1.5/20.2/20.6	2.1/18.5/19.1	0.8/9.3/10.1	0.8/10.1/10.4	1.9/16.4/17.4	2.3/18.1/18.7	2.0/14.4/15.1	2.1/17.0/17.7	1.8/10.6/11.4	-	5.3/20.5/19.7	3.9/16.2/17.0	1.6/13.5/14.6	1.6/8.1/9.1	2.1/17.2/18.1	0.4/3.5/4.1	2.1/14.4/15.4	1.6/7.9/8.6	2.1/12.6/13.0	1.9/15.5/17.2	0.5/2.1/2.9	1.6/13.8/14.6	0.3/2.2/3.1	0.9/12.5/13.2
kn	24.5/23.6/23.4	1.7/9.6/9.7	1.7/9.1/9.0	2.0/12.4/13.4	1.5/9.8/10.6	0.8/7.1/7.5	1.4/7.7/8.2	4.4/20.8/20.7	3.9/18.9/19.3	0.7/5.4/6.2	0.7/6.0/5.9	1.7/9.7/10.4	2.0/10.6/11.3	1.7/8.6/9.0	1.7/10.3/10.7	1.5/6.2/6.2	1.3/11.1/12.0	-	6.2/18.7/18.3	1.4/8.4/9.4	1.5/6.4/6.9	1.8/10.2/11.5	0.5/3.1/3.9	4.0/16.2/16.8	1.3/4.8/5.3	1.5/7.8/8.2	1.5/8.9/10.6	0.5/2.6/3.2	1.1/7.0/7.7	0.2/1.9/2.5	0.7/5.6/6.3
mr	26.1/24.4/24.5	1.9/10.3/10.8	1.7/9.5/9.6	2.3/13.0/14.0	1.6/10.1/10.6	1.0/8.1/8.4	1.4/8.1/8.9	5.3/23.8/23.5	4.4/20.5/21.2	0.6/5.9/6.6	0.7/6.4/6.5	2.0/11.0/11.4	2.1/11.7/13.0	1.9/9.0/9.7	1.8/10.5/11.4	1.6/6.5/7.0	1.4/12.3/12.5	7.6/23.1/22.8	-	1.4/8.7/10.1	1.5/6.5/7.4	1.9/12.0/13.0	0.4/2.9/3.4	4.1/18.9/19.7	1.4/5.2/6.1	1.8/8.2/8.7	1.8/9.3/10.9	0.4/2.2/3.0	1.2/7.6/8.6	0.2/2.0/3.1	0.7/7.0/7.3
mt	48.8/46.3/46.1	2.6/17.9/18.8	2.5/14.4/15.1	3.7/25.7/27.0	2.3/17.1/18.1	1.1/13.2/13.2	1.9/14.3/15.4	1.6/23.3/23.6	1.9/20.0/20.4	0.9/10.5/11.7	1.0/11.7/11.8	2.4/18.8/20.2	3.0/20.4/21.9	3.1/17.3/17.8	3.0/22.4/22.7	2.1/11.4/11.7	1.9/20.9/21.0	5.1/22.4/22.4	3.6/17.4/18.2	-	2.4/11.4/12.3	3.1/20.7/22.2	0.4/4.3/4.7	2.2/15.4/17.2	2.0/9.0/10.5	2.9/17.3/17.3	2.4/18.8/20.4	0.6/2.5/3.3	1.4/13.9/14.9	0.3/2.4/3.7	0.8/11.4/11.9
ha	10.4/19.5/19.3	0.5/8.4/9.1	0.5/7.4/8.1	1.2/11.1/12.2	0.6/8.4/9.4	0.2/6.4/7.3	0.5/6.9/7.3	0.4/12.1/12.6	0.4/10.2/11.5	0.2/5.0/5.8	0.1/5.3/5.3	1.1/9.0/9.5	1.0/9.9/10.8	1.5/7.4/8.1	0.8/9.2/10.0	0.8/6.1/6.5	0.4/10.2/10.9	1.4/13.6/14.8	0.6/9.4/11.1	0.9/8.5/9.5	-	1.1/10.8/11.4	0.1/2.2/3.0	0.6/8.4/9.9	0.7/5.3/5.9	0.6/7.7/7.9	1.3/8.0/9.3	0.1/1.3/2.3	1.0/6.5/7.6	0.1/1.5/2.3	0.1/6.0/6.4
af	48.7/45.9/45.0	3.9/17.8/18.5	3.1/15.3/16.0	5.2/21.5/23.2	3.1/14.1/15.2	1.7/12.3/12.9	2.1/13.4/13.7	2.1/21.0/21.4	3.3/18.3/18.9	1.8/9.3/10.2	1.4/10.5/10.7	3.9/19.4/20.3	4.1/20.3/22.1	4.1/13.6/14.7	4.2/18.6/19.3	3.2/9.7/10.4	3.9/18.7/18.9	5.4/20.3/19.7	4.4/15.9/17.3	2.7/17.1/19.0	2.5/9.3/10.2	-	0.5/3.7/4.6	3.0/14.4/15.9	2.1/9.4/11.1	3.1/14.1/14.7	2.5/15.0/16.5	0.6/1.9/2.8	2.7/11.6/13.0	0.3/2.3/3.2	1.0/10.2/10.6
am	21.6/19.2/19.5	2.6/8.1/8.1	2.1/7.4/7.7	4.1/9.8/10.8	2.8/7.9/8.3	1.8/5.9/6.5	1.5/6.0/6.3	2.9/13.6/13.4	3.9/12.4/12.9	1.7/4.5/4.9	1.2/4.7/5.0	2.5/7.4/8.3	2.8/8.4/9.3	3.1/6.7/7.3	2.6/8.5/8.9	1.7/4.5/5.0	3.9/9.9/10.0	7.1/15.6/14.6	4.1/11.2/11.9	1.5/6.9/8.1	1.6/4.9/5.9	2.4/9.0/9.7	-	3.0/9.9/11.0	1.4/3.5/4.4	1.9/5.9/6.3	1.9/6.7/7.6	0.6/1.4/2.0	1.8/5.7/6.4	0.8/3.0/3.8	1.0/5.1/5.6
gu	24.9/23.6/23.6	2.5/9.2/9.5	1.8/8.1/8.5	3.4/11.9/13.3	2.4/9.2/10.0	1.8/6.9/7.9	1.3/7.5/7.6	8.8/22.2/22.2	6.8/18.7/19.8	1.1/4.7/5.6	0.9/5.5/6.1	2.3/9.8/10.5	2.4/10.5/11.8	2.4/7.9/8.2	1.9/9.3/10.1	2.0/5.3/5.7	3.0/11.2/11.4	10.8/21.7/20.5	8.6/19.2/19.8	1.2/8.7/9.1	1.4/5.9/6.6	1.7/10.5/12.0	0.4/2.6/3.3	-	1.3/4.7/5.6	1.6/6.9/7.7	1.5/8.3/9.6	0.5/2.4/3.2	1.6/6.8/7.3	0.3/1.6/2.4	1.1/5.8/6.6
lb	21.6/21.9/24.6	4.4/15.6/17.6	2.4/11.2/12.4	3.2/13.3/15.9	2.3/10.1/11.8	1.2/7.0/8.6	1.8/8.5/9.7	1.7/13.7/15.1	1.8/11.7/13.9	1.0/4.6/5.6	0.6/5.4/6.3	2.5/11.6/14.0	2.7/12.5/15.2	2.8/9.3/10.7	2.5/11.1/13.1	2.1/6.4/7.2	2.2/11.0/13.0	3.4/13.5/15.6	2.4/10.2/11.8	1.7/9.0/10.9	1.8/5.8/7.0	2.4/12.9/15.7	0.3/2.1/2.6	1.6/9.4/11.4	-	2.0/8.8/10.1	1.9/8.8/11.7	0.4/1.3/1.7	1.7/6.9/8.5	0.2/1.2/2.0	0.7/6.2/7.4
oc	36.0/34.4/35.0	4.8/14.6/14.8	3.2/12.1/12.4	10.8/25.0/26.2	6.0/16.1/16.7	2.2/10.3/10.7	2.2/11.3/11.4	2.5/18.5/18.6	3.1/15.8/17.1	2.0/7.6/8.4	1.3/8.9/9.0	4.2/15.2/16.1	4.1/16.5/17.7	6.8/15.3/15.5	6.8/20.3/21.6	3.5/8.3/8.7	5.5/16.4/17.0	5.6/17.2/17.9	3.8/13.5/14.5	3.0/16.1/17.4	2.3/8.1/9.4	3.5/16.3/17.6	0.4/3.1/3.6	2.8/12.6/14.0	2.1/7.8/9.3	-	3.5/15.5/17.2	0.5/1.8/2.5	2.8/10.3/11.2	0.3/1.6/2.9	1.0/8.8/9.6
ro	31.1/29.7/31.2	3.9/15.2/16.0	2.9/12.6/12.8	7.9/23.8/24.7	4.6/16.2/16.4	1.9/12.0/12.9	2.5/13.0/13.8	4.4/19.9/19.9	4.2/17.1/18.1	1.7/8.4/9.3	1.1/9.8/9.8	3.7/16.1/17.2	3.8/17.1/18.1	5.3/16.5/17.0	4.5/18.6/19.5	3.3/10.0/10.6	3.1/18.3/18.9	8.0/19.4/20.2	5.0/14.3/15.6	2.6/15.4/16.8	1.7/8.4/9.3	2.7/15.9/17.5	0.3/3.4/3.7	3.0/13.6/15.1	1.8/7.9/9.3	3.0/14.4/14.4	-	0.4/2.2/3.1	1.8/12.0/13.0	0.3/2.0/2.8	0.8/9.9/10.9
sd	1.6/5.7/8.6	0.2/2.9/4.3	0.1/2.5/3.9	0.2/3.1/5.2	0.2/3.1/4.4	0.1/2.5/3.6	0.1/2.1/3.6	0.2/6.9/10.0	0.2/5.6/8.0	0.1/1.7/2.7	0.0/1.9/2.6	0.2/3.2/4.7	0.3/3.7/5.3	0.3/2.4/3.9	0.2/2.9/4.5	0.2/2.4/3.1	0.2/4.0/5.5	0.5/7.0/8.0	0.8/5.7/8.3	0.1/2.4/4.1	0.5/2.3/3.8	0.2/4.1/6.3	0.0/0.9/1.8	0.3/5.3/8.3	0.1/2.1/3.3	0.1/2.5/3.9	0.2/2.6/4.6	-	0.5/2.7/4.0	0.1/0.6/1.4	0.1/2.7/3.6
sr	31.4/29.0/29.0	3.4/13.6/14.9	1.8/11.2/11.5	5.3/18.3/19.4	3.2/13.2/14.1	2.4/12.4/13.1	1.7/13.0/13.4	4.1/17.9/18.0	4.8/15.4/15.9	1.8/7.7/8.5	1.3/8.5/8.8	2.4/14.3/15.1	3.5/15.4/16.5	3.8/12.2/12.9	3.1/14.4/15.3	3.2/9.3/9.8	4.7/18.5/18.9	7.7/17.8/17.8	5.6/14.0/15.0	1.8/12.6/13.5	1.4/7.4/7.9	2.3/14.6/15.7	0.4/3.1/3.7	2.9/12.1/13.4	1.3/6.4/7.8	2.4/10.9/11.1	1.9/13.3/15.0	0.4/1.6/2.5	-	0.3/1.9/2.7	0.9/10.7/11.1
ti	14.5/14.0/14.1	1.6/6.0/6.3	1.2/5.7/6.1	2.0/8.0/8.6	1.3/6.3/6.8	1.1/5.0/5.4	1.1/4.9/4.9	3.6/12.0/11.6	3.8/11.4/11.7	0.9/3.7/4.1	0.8/3.9/3.9	1.3/5.9/6.4	1.7/6.8/7.4	1.6/5.1/5.6	1.4/6.6/7.0	1.4/3.7/4.0	2.0/7.8/7.8	7.1/13.8/13.7	5.1/10.3/10.0	1.2/5.5/5.9	1.1/4.5/4.6	1.8/7.5/8.1	0.5/4.0/3.7	2.1/8.9/9.5	1.0/3.1/3.5	1.1/5.2/5.2	1.4/5.5/6.3	0.5/1.5/1.7	1.2/4.8/5.8	-	0.7/4.2/4.2
uk	26.8/26.2/27.0	3.3/12.8/13.4	1.8/10.9/11.0	5.8/17.0/17.9	3.6/12.3/13.1	5.3/16.3/16.1	2.0/11.5/12.3	3.4/16.4/16.8	4.0/14.5/15.0	1.9/6.7/7.6	1.2/7.9/7.9	3.2/13.0/13.9	2.8/14.5/15.3	3.3/11.3/12.2	3.2/13.4/14.0	3.1/9.4/9.8	5.8/17.4/17.6	5.8/16.7/16.9	4.1/12.8/13.8	1.8/10.5/11.9	1.2/6.3/7.5	2.2/13.2/14.1	0.3/2.5/3.1	2.5/11.5/12.7	1.3/6.1/7.0	2.3/9.7/10.1	2.1/11.7/13.3	0.4/1.6/2.3	1.8/10.8/12.1	0.2/1.5/2.4	-

Table 10: The SacreBLEU performance on EC30 for the baseline, boost-100, and boost-all models in the one-tag fashion, respectively. 970 directions of results are shown, including English-centric and zero-shot ones. Overall, in 869 out of 870 zero-shot directions, the boost-100 model achieves better performance compared to the baseline. Moreover, in 423 out of 870 zero-shot directions, the gain exceeds 3.0.

COMET	en	de	nl	fr	es	ru	cs	hi	bn	ar	he	sv	da	it	pt	pl	bg	kn	mr	mt	ha	af	am	gu	lb	oc	ro	sd	sr	ti	uk
en	-	82.5/82.1/81.4	84.0/83.8/83.6	82.5/81.6/81.3	81.6/80.9/80.8	81.8/81.1/80.9	84.4/83.6/83.1	77.6/78.0/77.8	84.3/84.5/84.5	78.9/78.7/78.8	82.9/83.3/83.4	84.9/83.9/83.6	85.2/84.2/83.6	83.5/82.7/82.5	84.5/84.1/83.5	81.8/81.4/81.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
de	85.8/85.5/85.2	-	72.5/76.5/76.5	63.7/68.3/69.1	64.0/69.4/69.8	58.6/70.1/70.7	57.8/72.6/73.9	61.4/66.6/66.7	70.2/77.2/77.0	58.1/64.9/66.3	57.2/67.4/68.6	70.1/73.5/74.3	69.6/73.4/74.2	66.6/69.6/71.2	67.8/70.8/71.7	63.6/69.6/70.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
nl	84.2/84.0/83.8	69.2/71.7/72.4	-	63.7/67.1/67.6	65.1/68.2/69.0	60.4/69.0/69.9	57.6/70.3/71.7	62.7/66.5/66.6	72.6/77.5/77.5	59.7/64.3/65.6	57.8/66.8/68.7	68.8/71.4/72.1	68.9/72.2/72.5	66.7/69.1/70.0	67.3/69.8/70.5	63.3/69.4/69.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
fr	85.5/84.9/84.9	65.8/69.1/69.6	70.5/72.7/73.8	-	73.3/74.5/74.3	65.3/71.6/72.3	60.3/71.3/73.0	61.4/67.4/67.6	70.3/76.9/77.1	65.0/68.7/69.4	61.8/70.3/71.2	68.4/70.9/71.4	67.6/71.3/72.1	75.5/76.2/76.3	75.1/75.4/76.2	67.0/70.9/71.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
es	82.9/82.8/82.5	61.7/65.9/66.9	66.0/70.0/71.1	68.9/70.8/71.1	-	62.8/69.7/70.7	56.5/69.4/70.3	61.3/66.8/67.5	70.4/77.5/77.2	62.7/66.6/68.2	59.1/67.2/68.8	66.1/68.3/69.7	65.0/69.4/70.0	73.5/74.5/74.9	78.3/77.5/77.6	65.0/68.7/69.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ru	81.6/81.2/80.7	61.2/65.7/66.5	64.7/69.4/70.0	62.5/65.7/66.0	64.6/67.8/68.4	-	62.6/71.8/72.8	59.8/64.9/64.6	70.7/75.7/75.5	60.2/64.0/65.5	57.8/65.8/67.2	64.5/67.6/68.4	63.2/68.3/68.9	65.6/68.4/69.1	66.1/69.3/69.8	66.4/70.8/71.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
cs	84.5/84.1/84.0	64.5/69.5/70.7	67.5/72.3/73.1	63.3/68.0/68.6	65.3/69.3/70.0	65.3/72.8/73.2	-	60.6/66.1/66.1	69.4/75.6/76.1	59.5/64.1/65.6	57.6/67.0/68.3	66.1/70.3/71.0	65.1/70.2/70.8	66.2/69.7/70.9	66.5/70.1/70.9	70.2/72.4/73.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
hi	84.7/84.5/84.1	57.5/65.6/66.5	62.0/70.0/70.5	62.9/66.9/67.6	65.2/69.3/69.7	50.2/68.8/69.7	56.5/68.5/69.8	-	69.1/81.8/81.6	51.3/64.9/66.3	55.5/66.5/67.7	65.7/69.6/70.2	66.6/70.0/70.6	62.4/67.6/68.4	65.2/70.9/71.2	57.3/66.3/67.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
bn	81.8/81.5/81.3	53.4/62.4/63.3	59.2/66.7/68.2	59.0/63.3/63.5	61.5/66.0/66.5	46.9/66.1/67.5	52.9/64.8/65.9	52.0/69.0/69.7	-	48.0/63.0/64.1	53.1/64.3/65.8	62.2/66.3/67.1	62.9/66.8/67.6	58.6/64.8/65.9	60.9/67.2/67.9	53.9/63.4/64.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ar	78.9/78.3/78.3	54.7/58.9/60.0	59.4/63.5/64.2	59.3/62.0/62.3	61.4/64.1/64.7	56.3/64.1/64.9	53.8/61.5/63.3	56.2/62.1/61.9	66.8/73.8/75.0	-	56.0/63.6/65.2	59.7/63.1/63.7	58.6/62.9/64.0	60.3/63.4/64.8	61.5/64.6/66.0	56.3/61.1/61.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
he	83.2/82.7/82.2	57.2/62.6/63.6	62.4/67.4/68.2	62.1/64.5/65.1	63.3/66.0/66.9	55.6/65.9/67.0	55.3/64.5/66.1	56.1/63.2/63.8	66.7/75.2/75.3	59.6/64.0/65.8	-	61.4/65.1/66.5	61.3/65.2/66.3	62.3/65.3/66.8	63.8/66.6/67.8	57.1/63.0/63.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
sv	85.1/84.6/84.6	65.9/69.2/70.3	68.3/71.8/72.8	63.0/66.1/67.2	64.4/68.6/69.0	62.4/68.7/69.2	59.0/70.0/71.5	63.4/67.3/67.6	72.4/77.1/77.7	60.4/64.5/66.3	59.4/66.4/68.3	-	76.7/76.7/77.1	66.4/68.7/69.8	67.2/70.0/71.0	64.4/67.5/68.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
da	85.3/84.8/84.7	66.1/68.9/70.3	69.8/72.4/73.2	62.3/65.8/66.8	64.4/67.7/69.1	63.7/69.3/70.0	59.0/68.5/71.2	64.1/67.5/68.1	73.5/77.5/78.2	60.7/64.3/66.3	59.6/66.0/68.5	77.4/76.5/76.9	-	66.3/68.3/69.6	67.1/69.5/70.8	64.0/67.5/68.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
it	84.0/83.6/83.5	64.3/67.3/68.0	68.3/71.4/71.9	70.9/72.4/72.7	73.3/74.9/74.7	62.3/70.7/71.3	55.5/70.1/71.5	59.5/66.0/66.0	69.8/76.3/76.7	61.4/66.2/67.5	59.0/67.5/68.8	66.4/69.2/70.3	66.3/70.2/70.8	-	74.2/75.4/75.8	66.0/69.1/69.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
pt	85.1/84.8/84.7	63.8/67.2/67.8	67.7/71.4/71.6	69.9/72.0/72.3	77.5/77.3/76.8	64.3/69.6/70.0	55.8/68.8/71.0	62.2/68.0/68.0	71.7/77.6/78.0	61.7/65.9/67.0	59.7/67.3/69.1	67.8/70.3/70.7	66.7/70.1/71.2	73.8/74.9/75.3	-	65.2/67.5/68.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
pl	80.6/80.3/80.1	61.1/65.9/66.8	63.8/69.7/70.7	60.7/65.4/66.0	63.3/67.6/68.3	63.5/71.5/72.1	66.8/72.2/73.5	58.7/64.3/64.1	67.3/74.2/74.6	57.3/62.4/64.6	54.0/64.5/65.7	64.5/68.0/68.4	63.8/68.0/68.7	65.3/68.2/69.4	65.0/68.5/69.5	-	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
bg	84.9/84.5/84.4	64.9/68.5/68.9	68.2/72.2/72.4	66.1/68.8/69.3	68.3/70.5/70.6	75.8/78.3/78.5	63.8/72.8/73.6	61.4/65.9/66.1	71.2/76.4/76.8	61.7/65.7/66.9	60.3/68.0/69.5	68.4/70.2/70.8	67.3/70.7/71.7	69.4/71.0/72.0	69.6/71.3/72.6	68.3/70.9/71.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
kn	79.1/79.0/78.6	51.9/60.4/61.3	56.7/65.1/65.7	57.1/61.0/61.2	59.7/63.7/64.5	46.8/64.5/65.3	51.1/62.2/63.8	53.6/67.4/67.7	68.2/79.1/79.5	47.5/61.4/63.2	51.0/62.7/63.5	60.9/65.4/65.6	61.2/65.2/66.0	57.3/62.8/63.5	60.3/65.2/66.0	52.5/61.3/62.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
mr	80.9/80.8/80.5	53.3/61.4/61.8	57.4/65.6/66.5	59.0/61.9/62.7	61.2/65.0/65.6	45.6/65.0/66.5	51.4/63.6/65.3	53.9/71.2/71.7	65.6/80.4/80.8	47.7/61.0/62.9	51.8/62.5/63.9	62.8/66.1/66.9	62.9/66.7/67.5	58.6/63.6/65.0	62.1/66.8/67.6	53.4/62.5/64.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
mt	79.5/79.4/79.0	53.8/60.6/62.2	56.9/64.5/65.3	58.3/62.6/63.2	60.4/64.2/65.1	50.3/64.0/64.8	45.9/63.5/64.9	55.2/64.8/65.3	63.5/76.0/76.5	53.9/60.8/62.8	48.9/62.2/64.2	59.4/64.1/64.8	59.3/64.4/65.7	62.4/65.1/66.4	62.6/66.3/67.5	54.3/61.3/62.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ha	51.9/66.8/67.1	37.5/47.6/49.5	40.0/52.5/54.7	37.7/49.1/51.1	40.4/52.6/54.6	34.7/52.5/55.1	32.3/49.6/53.8	42.1/57.6/58.7	50.6/69.5/71.0	35.3/50.7/55.3	34.2/51.1/54.3	40.2/51.4/54.2	38.8/52.5/55.0	40.5/51.7/53.8	41.0/53.2/55.8	33.3/51.1/53.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
af	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
am	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
gu	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
lb	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
oc	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ro	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
sd	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
sr	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ti	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
uk	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 11: The COMET performance on EC30 for the baseline, boost-100, and boost-all models in the one-tag fashion, respectively. 970 directions of results are shown, including English-centric and zero-shot ones. Overall, in 866 out of 870 zero-shot directions, the boost-100 model achieves better performance compared to the baseline. Moreover, in 501 out of 870 zero-shot directions, the gain exceeds 5.0.

	Model	H-H	H-M	H-L	M-H	M-M	M-L	L-H	L-M	L-L	AVG
Two-Tag	Baseline	1.5	2.2	1.3	1.8	2.4	1.3	2.6	3.1	1.3	2.0
	Boost-100	13.4	14.2	9.0	13.8	14.2	9.2	10.8	11.2	6.8	11.4
	Boost-All	14.0	14.9	10.0	14.4	14.9	10.1	11.6	12.1	7.8	12.2
	Δ-100	+11.9	+12.0	+7.7	+12.0	+11.8	+7.9	+8.2	+8.1	+5.5	+9.4
	Δ-All	+12.5	+12.7	+8.7	+12.6	+12.5	+8.8	+9.0	+9.0	+6.5	+10.2
One-Tag	Baseline	8.4	9.0	4.6	8.8	8.9	4.7	6.4	6.7	3.1	6.7
	Boost-100	12.4	12.8	8.0	12.7	12.6	8.1	9.5	9.7	5.7	10.2
	Boost-All	12.9	13.4	9.0	13.2	13.3	9.1	10.4	10.6	6.8	10.9
	Δ-100	+4.0	+3.8	+3.4	+3.9	+3.7	+3.4	+3.1	+3.0	+2.6	+3.5
	Δ-All	+4.5	+4.4	+4.4	+4.4	+4.4	+4.4	+4.0	+3.9	+3.7	+4.2

Table 12: Zero-shot performance (SacreBLEU) on the EC30 dataset (61M sentence pairs) with two language tag strategies, grouped by **H**igh-, **M**edium, and **L**ow-resource, respectively. Δ-100 and Δ-All mean the corresponding performance changes compared with the baselines.

	Model	High		Medium		Low		Overall		
		EN-X	X-EN	EN-X	X-EN	EN-X	X-EN	EN-X	X-EN	AVG
Two-Tag	Baseline	28.0	31.4	29.5	31.4	18.7	25.8	25.4	29.5	27.5
	Boost-100	27.3	29.9	29.0	31.1	18.8	25.0	25.0	28.7	26.9
	Boost-All	27.1	29.7	28.7	31.1	19.3	25.8	25.0	28.9	27.0
	Δ-100	-0.7	-1.5	-0.5	-0.3	+0.1	-0.8	-0.4	-0.8	-0.6
	Δ-All	-0.9	-1.7	-0.8	-0.3	+0.6	0.0	-0.4	-0.6	-0.5
One-Tag	Baseline	28.4	30.9	29.8	30.9	19.5	25.3	25.9	29.0	27.5
	Boost-100	27.6	30.5	29.6	31.4	19.4	25.5	25.5	29.1	27.3
	Boost-All	27.0	30.0	28.7	31.2	19.6	25.5	25.1	28.9	27.0
	Δ-100	-0.8	-0.4	-0.2	+0.5	-0.1	+0.2	-0.4	+0.1	-0.2
	Δ-All	-1.4	-0.9	-1.1	+0.3	+0.1	+0.2	-0.8	-0.1	-0.5

Table 13: English-centric performance (SacreBLEU) on the EC30 dataset (61M sentence pairs). EN-X and X-EN denote the average out-of- and into-English translation performance of each resource group, respectively.

	Model	H-H	H-M	H-L	M-H	M-M	M-L	L-H	L-M	L-L	AVG
Two-Tag	Baseline	51.0	52.4	47.3	48.6	49.9	45.0	46.4	49.3	43.6	48.2
	Boost-100	70.6	71.2	64.0	69.1	69.7	63.1	63.1	64.8	58.7	66.0
	Boost-All	71.3	71.8	65.7	69.9	70.4	64.8	64.4	66.0	60.9	67.3
	Δ-100	+19.6	+18.8	+16.7	+20.5	+19.8	+18.1	+16.7	+15.5	+15.1	+17.8
	Δ-All	+20.3	+19.4	+18.4	+21.3	+20.5	+19.8	+18.0	+16.7	+17.3	+19.1
One-Tag	Baseline	61.2	62.6	54.4	58.8	60.2	52.2	51.9	54.9	47.8	56.0
	Boost-100	68.1	68.8	61.7	66.5	67.1	60.5	60.2	62.1	55.8	63.4
	Boost-All	68.9	69.5	63.6	67.6	68.0	62.7	62.1	63.7	58.7	65.0
	Δ-100	+6.9	+6.2	+7.3	+7.7	+6.9	+8.3	+8.3	+7.2	+8.0	+7.4
	Δ-All	+7.7	+6.9	+9.2	+8.8	+7.8	+10.5	+10.2	+8.8	+10.9	+9.0

Table 14: Zero-shot performance (COMET) on the EC30 dataset (61M sentence pairs) with two language tag strategies, grouped by **H**igh-, **M**edium, and **L**ow-resource, respectively. Δ-100 and Δ-All mean the corresponding performance changes compared with the baselines.

	Model	High		Medium		Low		Overall		
		EN-X	X-EN	EN-X	X-EN	EN-X	X-EN	EN-X	X-EN	AVG
Two-Tag	Baseline	82.3	83.8	81.0	80.0	73.7	73.3	79.0	79.0	79.0
	Boost-100	82.0	83.4	81.1	81.3	73.7	74.4	78.9	79.7	79.3
	Boost-All	82.1	83.0	81.0	81.1	74.3	75.2	79.1	79.8	79.5
	Δ-100	-0.3	-0.4	+0.1	+1.3	0.0	+1.1	-0.1	+0.7	+0.3
	Δ-All	-0.2	-0.8	0.0	+1.1	+0.6	+1.9	+0.1	+0.8	+0.5
One-Tag	Baseline	82.0	83.3	80.6	79.6	73.2	72.6	78.6	78.5	78.5
	Boost-100	81.8	83.0	80.8	80.9	73.7	74.2	78.8	79.4	79.1
	Boost-All	81.6	82.7	80.6	80.7	73.9	74.7	78.7	79.4	79.0
	Δ-100	-0.2	-0.3	+0.2	+1.3	+0.5	+1.6	+0.2	+0.9	+0.6
	Δ-All	-0.4	-0.6	0.0	+1.1	+0.7	+2.1	+0.1	+0.9	+0.5

Table 15: English-centric performance (COMET) on the EC30 dataset (61M sentence pairs). EN-X and X-EN denote the average out-of- and into-English translation performance of each resource group, respectively.

	Model	H-H	H-M	H-L	M-H	M-M	M-L	L-H	L-M	L-L	AVG
Two-Tag	M2M100	40.0	35.6	24.8	35.0	30.3	21.8	33.1	29.9	20.3	30.1
	Baseline	11.0	14.8	10.6	11.3	14.9	10.5	13.7	17.4	10.9	12.8
	Boost-100	37.9	38.9	32.2	37.6	38.0	31.7	34.5	35.2	28.6	35.0
	Boost-All	38.6	39.6	33.5	38.3	38.6	33.0	35.6	36.3	30.3	36.0
	Δ_{M2M}-100	-2.1	+3.3	+7.4	+2.6	+7.7	+9.9	+1.4	+5.3	+8.3	+4.9
	Δ_{M2M}-All	-1.4	+4.0	+8.7	+3.3	+8.3	+11.2	+2.5	+6.4	+10.0	+5.9

Table 16: The performance (ChrF++) of our methods compared to M2M100 (418M parameters) in the involved directions of EC30. Note that some scores slightly differ from those presented in Table 1. This disparity arises because M2M100 lacks support for the medium-resource language *mt* and the low-resource language *ti*. To ensure a fair comparison, we evaluate in 756 zero-shot directions, where these two languages are excluded. Δ_{M2M}-100 and Δ_{M2M}-All denote the performance gap between M2M100 and our two boosted models (Boost-100 and Boost-All).

Data Size	1	2	5	10	25	50	100
AVG (ChrF++)	36.0	36.9	38.0	38.6	38.9	39.2	39.3
STDEV	0.5	0.5	0.2	0.2	0.2	0.1	0.1

Table 17: We report the mean and variance of the overall zero-shot performance for each step in Figure 4 after running with three random seeds. Across sample sizes ranging from 1 to 100, we note a consistently small standard deviation. Furthermore, with increasing sample sizes, the standard deviation diminishes further, indicating a stable performance across observations.

Metric	Model	Within Family					Out of Germanic				Into Germanic			
		Germanic	Romance	Slavic	Aryan	Asiatic	Romance	Slavic	Aryan	Asiatic	Romance	Slavic	Aryan	Asiatic
ChrF++	Baseline	20.8	23.5	12.1	8.0	7.2	21.0	10.5	5.8	7.3	20.9	20.0	16.8	17.8
	Boost-All	44.6	**45.6**	**39.2**	**34.2**	**27.4**	**42.2**	**38.0**	**34.1**	**30.9**	**42.6**	**40.6**	**34.8**	**36.4**
	Boost-Germanic	**45.0**	44.5	38.5	30.3	25.0	41.3	37.4	32.4	29.4	42.4	40.5	33.2	35.9
Comet	Baseline	54.3	58.5	50.4	45.9	41.6	56.7	47.6	39.5	41.3	54.9	55.4	52.8	50.0
	Boost-All	68.2	**71.6**	**74.0**	**71.2**	**66.2**	**66.3**	**69.2**	**69.5**	**68.2**	**67.3**	**67.3**	**63.4**	**61.1**
	Boost-Germanic	**68.5**	70.8	72.8	68.0	63.3	65.8	68.3	68.7	66.8	66.9	66.9	61.8	60.1
SacreBLEU	Baseline	2.8	3.7	2.0	3.7	0.8	2.7	1.6	2.5	1.1	2.6	2.2	1.6	1.7
	Boost-All	17.5	**18.7**	**13.5**	**15.9**	**6.5**	**16.4**	**12.8**	**15.3**	**8.6**	**15.1**	**13.8**	**9.7**	**11.3**
	Boost-Germanic	**17.8**	17.4	12.7	14.3	5.3	15.3	12.2	14.1	7.6	14.8	13.6	9.1	10.9

Table 18: Zero-shot performance on the EC30 dataset. Boost-All means fine-tuning using all of the NTREX data, while Boost-Germanic means partially fine-tuning using Germanic languages. The results are in three groups: 1) "Within Family" shows the performance within a specific language group, where both the source and target languages belong. 2) "Out of Germanic" shows the average performance that is translated out of Germanic languages, e.g., from Germanic to Romance. and 3) "Into Germanic" shows the average performance that is translated into Germanic languages, e.g., from Romance to Germanic.

ChrF++	en	da	de	es	fi	fr	it	nl
en	-	55.5/55.1/53.6	50.6/51.0/49.0	46.2/46.3/44.8	44.2/43.3/42.8	57.4/56.8/55.3	47.9/47.0/46.3	46.5/46.8/45.3
da	57.2/57.1/55.4	-	28.3/44.3/47.2	24.8/39.9/42.3	22.4/38.5/41.8	30.5/45.6/50.4	26.8/39.5/43.3	23.4/41.9/44.2
de	53.4/53.1/52.0	28.5/44.9/48.7	-	24.6/39.3/41.9	21.4/36.6/40.5	30.2/44.4/49.0	26.9/39.1/42.6	23.4/41.4/43.7
es	48.0/47.8/46.8	26.5/39.7/43.3	25.9/38.3/40.8	-	20.9/34.8/38.1	30.5/44.0/47.4	28.3/40.6/43.2	22.6/38.2/40.7
fi	45.9/45.8/45.4	26.7/39.0/42.7	26.8/36.8/40.6	25.2/35.9/38.8	-	28.4/39.4/44.1	26.6/35.2/39.3	22.5/36.4/39.6
fr	56.0/55.3/54.0	27.3/43.8/47.4	26.8/42.7/45.7	25.7/42.4/44.5	20.7/37.6/41.2	-	28.4/43.8/46.3	23.0/41.2/43.5
it	50.2/49.6/48.9	26.4/40.8/44.4	25.8/39.8/43.0	26.5/41.6/43.9	21.1/35.1/39.3	31.2/46.4/49.8	-	22.9/39.0/41.9
nl	48.2/47.9/47.0	27.0/40.8/43.6	26.7/39.9/42.0	24.5/37.3/39.5	20.7/34.8/37.8	28.4/41.3/44.5	26.1/36.8/39.5	-

Table 19: The detailed ChrF++ results for the baseline, boosted, and upper bound models on the Europarl-8 dataset are presented, encompassing 14 English-centric directions and 42 zero-shot directions.

COMET	en	da	de	es	fi	fr	it	nl
en	-	79.5/78.8/76.9	72.8/72.5/70.4	75.1/74.9/71.6	77.5/76.6/75.7	75.3/73.8/72.0	76.1/74.9/72.8	75.4/75.0/73.0
da	78.6/78.7/77.6	-	49.7/64.8/71.8	55.3/65.3/71.4	50.4/68.5/75.6	52.7/61.6/70.3	53.1/64.6/72.4	52.4/68.1/73.5
de	76.6/76.6/75.8	56.1/71.6/76.1	-	54.2/63.6/69.7	49.0/66.2/73.9	51.4/61.7/68.6	52.6/64.2/71.4	51.0/68.5/73.7
es	75.5/75.5/74.6	54.8/66.6/73.7	46.9/60.7/67.3	-	48.2/66.5/73.4	53.4/65.5/72.1	54.5/69.6/75.8	50.3/64.6/70.7
fi	75.2/74.3/75.0	55.6/66.5/73.6	47.9/59.3/67.4	54.5/62.5/69.2	-	51.7/59.5/68.6	52.7/63.0/69.9	51.7/63.1/69.9
fr	79.1/79.1/77.6	57.4/69.6/75.5	48.6/63.7/70.6	57.9/70.9/75.2	50.3/68.9/75.6	-	56.3/72.5/77.2	52.7/68.4/73.2
it	76.6/76.6/75.8	55.8/67.8/74.2	47.4/61.3/69.2	56.6/70.4/74.8	48.8/66.6/74.3	54.8/67.3/73.4	-	50.8/65.6/71.8
nl	75.7/75.6/75.0	55.6/69.2/75.0	48.6/64.1/70.2	54.0/64.2/69.5	48.3/66.0/72.9	51.4/61.0/68.4	51.7/64.1/70.6	-

Table 20: The detailed COMET results for the baseline, boosted, and upper bound models on the Europarl-8 dataset are presented, encompassing 14 English-centric directions and 42 zero-shot directions.

SacreBLEU	en	da	de	es	fi	fr	it	nl
en	-	30.4/30.4/28.1	23.0/23.3/21.6	18.9/19.2/17.9	13.7/12.8/12.9	32.6/30.8/30.6	19.9/18.9/18.7	17.6/18.0/16.9
da	31.1/31.0/28.3	-	6.4/16.4/19.5	4.2/13.7/15.9	3.0/9.7/12.0	7.9/18.4/24.3	4.9/12.5/15.8	3.3/13.4/15.1
de	26.5/26.5/24.3	7.0/18.9/22.8	-	4.3/13.0/15.7	2.8/8.6/11.1	7.6/17.6/22.9	5.4/12.5/15.1	3.5/13.3/15.3
es	18.4/18.5/17.4	5.0/13.0/15.9	4.2/10.2/12.1	-	2.0/6.5/8.2	6.9/15.1/20.3	5.5/11.8/15.1	2.7/9.5/11.4
fi	18.5/18.4/17.9	5.7/13.7/16.8	4.8/10.1/13.1	4.5/10.5/12.7	-	6.1/13.5/18.0	4.5/9.2/12.7	2.8/9.0/11.3
fr	28.8/28.4/26.2	5.8/17.3/20.6	5.4/14.2/17.5	4.6/15.8/17.8	2.4/9.3/11.5	-	6.1/16.0/18.6	3.2/12.5/14.5
it	20.7/20.6/19.4	5.1/13.8/16.7	4.5/11.4/14.3	5.0/14.2/17.0	2.3/7.1/9.5	7.9/18.4/22.8	-	2.9/10.0/12.1
nl	20.7/20.5/19.2	5.6/14.6/17.0	5.1/12.2/13.7	4.2/11.4/13.3	2.4/7.4/8.9	6.6/14.7/18.0	4.3/10.3/12.1	-

Table 21: The detailed SacreBLEU results for the baseline, boosted, and upper bound models on the Europarl-8 dataset are presented, encompassing 14 English-centric directions and 42 zero-shot directions.

Figure 6: Illustration of number pairs.

Source	*Target*
961 271 4 137 146 37 124 498 28 73	961 271 4 137 146 37 124 498 28 73
323 135 42 121 324 499 17 11 37 33	323 135 42 121 324 499 17 11 37 33
713 93 331 225 51 7 375 223 12 192	713 93 331 225 51 7 375 223 12 192
...	...