

## A AMD AND DOWNSTREAM PERFORMANCE

We evaluate the relationship between AMD and downstream performance on NMT tasks. We focus on BPE and sample vocabularies with merging sizes varying from 1K to 50K. The relation is evaluated on 12 language-pairs. As shown in Table 1, 42.8% percent of language-pairs with the highest AMD achieve the best BLEU scores. 35.7% percent of language-pairs with the highest AMD achieve almost the best BLEU scores (gap less than 0.3).

Table 1: AMD is related with NMT performance. 42.8% percent language-pairs with the highest AMD achieve the best BLEU scores. 35.7% percent of language-pairs with the highest AMD achieve almost the best BLEU scores (gap less than 0.3).

		1K	5K	10K	30K	50K
WMT14 En-De	AMD	-3.33E-3	-3.66E-4	<b>9.89E-5</b>	6.76E-6	3.03E-6
	BLEU	29.30	29.61	<b>29.84</b>	29.31	29.09
WMT16 En-Ro	AMD	-3.19E-4	4.23E-5	2.68E-5	5.65E-6	1.94E-6
	BLEU	36.5	36.6	36.6	<b>36.9</b>	36.4
En-Es	AMD	-3.89E-04	1.20E-05	<b>1.75E-05</b>	5.32E-06	2.40E-06
	BLEU	40.80	<b>40.98</b>	40.78	39.97	39.75
En-PTbr	AMD	-3.82E-04	1.20E-05	<b>1.72E-05</b>	5.18E-06	2.39E-06
	BLEU	<b>41.32</b>	41.24	41.10	40.32	40.01
En-Fr	AMD	-3.52E-04	<b>2.03E-05</b>	1.82E-05	5.07E-06	2.20E-06
	BLEU	44.40	44.91	<b>44.98</b>	43.98	43.85
En-Ru	AMD	-4.46E-04	-8.74E-06	<b>1.24E-05</b>	5.13E-06	2.73E-06
	BLEU	19.88	20.55	<b>20.90</b>	20.11	19.86
En-He	AMD	-6.04E-04	-3.32E-05	<b>3.61E-06</b>	3.03E-06	1.79E-06
	BLEU	29.59	<b>29.70</b>	29.41	28.76	28.64
En-Ar	AMD	-5.55E-04	-3.42E-05	<b>3.86E-06</b>	3.26E-06	1.91E-06
	BLEU	18.82	<b>19.21</b>	18.83	18.25	17.92
En-Ko	AMD	<b>-3.28E-04</b>	-3.35E-05	-6.45E-06	-7.39E-07	-4.34E-08
	BLEU	10.67	10.82	<b>10.94</b>	10.40	10.21
En-It	AMD	-4.01E-04	2.35E-06	<b>1.45E-05</b>	5.86E-06	3.46E-06
	BLEU	37.07	<b>37.50</b>	37.43	36.88	36.56
En-Nl	AMD	-4.00E-04	1.59E-05	<b>1.83E-05</b>	5.38E-06	2.49E-06
	BLEU	<b>34.21</b>	34.05	33.84	33.42	33.38
En-Ro	AMD	-4.53E-04	2.36E-06	<b>1.48E-05</b>	5.18E-06	2.58E-06
	BLEU	29.49	29.46	<b>29.65</b>	28.50	28.66
En-Tr	AMD	-3.94E-04	6.11E-06	<b>1.57E-05</b>	5.27E-06	2.64E-06
	BLEU	20.21	<b>20.30</b>	19.75	19.12	19.01
En-De	AMD	-3.82E-04	1.10E-05	<b>1.80E-05</b>	6.04E-06	2.93E-06
	BLEU	31.38	31.36	<b>31.50</b>	30.28	30.11

## B PROOFS FOR EQ.10

$$\begin{aligned}
\sum_{i \in \mathbb{T}} P(i) \log P(i) &= \sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{C}} P(i, j) \log P(i) \\
&= \sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{C}} P(i, j) \log P(i, j) \cdot \frac{P(i)}{P(i, j)} \\
&= \sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{C}} P(i, j) \log P(i, j) + \sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{C}} P(i, j) \log \frac{P(i)}{P(i, j)} \\
&= \underbrace{\sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{C}} P(i, j) \log P(i, j)}_{\mathcal{L}_1} + \underbrace{\sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{C}} P(i, j) (-\log P(j|i))}_{\mathcal{L}_2}.
\end{aligned}$$

□

---

## C ISO LANGUAGE CODE

The language names and their corresponding language codes according to ISO 639-1 standard are listed in Table 2.

Table 2: ISO Language code

Es	Spanish
Ptbr	Portuguese
Fr	French
Ru	Russian
He	Hebrew
Ar	Arabic
Ko	Korean
It	Italian
Nl	Dutch
Tr	Turkish
De	German
Ro	Romanian