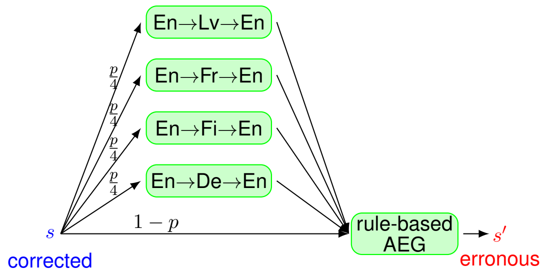

Figure 1: Integration of Our Method and RTT

p	BEA-19 test	CoNLL 14	JFLEG test
0.0	67.32 / 69.42	60.60 / 62.25	60.12 / 60.69
0.1	67.52 / 69.28	60.73 / 62.78	60.21 / 60.84
0.2	67.35 / **69.83**	60.58 / 62.69	60.19 / 60.69
0.3	66.60 / 68.58	60.37 / 61.78	**60.38 / 60.91**
0.4	**68.00** / 69.76	60.48 / 62.13	59.90 / 60.46
0.5	67.92 / 69.71	60.55 / 62.02	60.15 / 60.74
0.6	67.28 / 69.02	60.75 / 61.65	60.20 / 60.60
0.7	67.80 / 69.38	**61.29 / 63.47**	60.26 / 60.75
0.8	67.94 / 69.53	60.22 / 61.81	60.32 / 60.74
0.9	67.09 / 69.25	60.49 / 62.95	60.04 / 60.60
1.0	67.65 / 69.08	60.63 / 61.44	60.21 / 60.57

Table 1: Integration of our method and RTT.

A Integration of Our Method and RTT

We further investigate if combination of our method and RTT is beneficial. We integrate them by applying RTT at the probability of p, in which RTT'ed sentence is selected with an equal probability, then conducting our method. Figure 1 shows the integrated procedure.

Table 1 shows the results. The optimal p varies for each evaluation dataset, and it is difficult to say if the integration is beneficial. There might be a good method to combine our method and MT-based methods. We plan to research this in future work.

B Effectiveness of BPE-dropout

All the experiments in our paper use BPE-dropout of $p = 0.1$, which is recommended in the original paper. We investigate the effectiveness of subword regularization with changing p.

Table 2 shows the result for baseline models. The optimal p is 0.1. BPE-dropout improves performance as subword reguralization.

Table 3 shows the result for artificial+target settings using 16M augmented data. BPE-dropout applies to both pre-training and fine-tuning data at the same dropout probability. The optimal p varies for each evaluation dataset. Furthermore, we can observe that BPE-dropout improves even pre-trained and fine-tuned models in BEA-19 test and JFLEG test dataset.

p	BEA-19 test	CoNLL 14	JFLEG test
0.0	55.21 / 61.21	49.68 / 53.45	53.73 / 54.26
0.05	58.87 / **63.56**	53.26 / 56.51	57.19 / 57.78
0.1	**59.09 / 63.56**	**54.03 / 56.79**	**57.33 / 58.15**
0.15	58.62 / 63.27	53.31 / 56.36	57.25 / 57.73
0.20	57.59 / 62.60	53.54 / 56.42	57.24 / 57.82

Table 2: The effect of BPE-dropout for target-only settings.

p	BEA-19 test	CoNLL 14	JFLEG test
0.0	66.51 / 69.34	60.36 / **63.14**	59.55 / 60.04
0.05	67.19 / **70.32**	59.68 / 62.73	60.41 / **61.03**
0.1	**67.32** / 69.42	**60.60** / 62.25	60.12 / 60.69
0.15	66.90 / 69.26	60.53 / 61.97	60.31 / 60.71
0.20	66.96 / 69.34	60.32 / 62.28	**60.45** / 60.95

Table 3: The effect of BPE-dropout for artificial+target settings.