

Appendix

Full development set results for En-De translation and En-Zh translation.

Order	English → German			English → Chinese	
	τ	BLEU (+EOS)	BLEU (+EOS)	BLEU (+EOS)	BLEU (+EOS)
	+Parallel			+Parallel	
		Stern et al. (2019)		This Work	
Uniform	∞	22.39 (25.58)	24.31 (24.91)	28.6 (31.8)	30.4 (31.9)
Binary Tree	0.5	24.49 (25.55)	25.33 (25.70)	29.3 (31.6)	31.3 (31.9)
	1.0	24.36 (25.43)	25.43 (25.76)	29.6 (32.0)	31.4 (32.2)
	2.0	24.59 (25.80)	25.33 (25.80)	29.1 (32.2)	31.4 (32.3)
		This Work			
Random	0.5	23.82 (24.87)	23.97 (24.20)	28.5 (30.6)	29.4 (30.2)
	1.0	24.03 (25.46)	24.58 (24.82)	28.6 (31.1)	30.0 (31.0)
	2.0	24.00 (25.41)	24.68 (25.07)	28.9 (31.7)	30.4 (31.6)
L2R (Left-Aligned)	0.5	21.19 (24.46)	21.40 (21.57)	24.5 (30.0)	25.7 (28.3)
	1.0	21.36 (24.02)	20.84 (21.25)	24.8 (29.8)	25.2 (27.8)
	2.0	21.78 (24.21)	20.56 (21.11)	25.8 (29.8)	24.9 (27.6)
L2R (Right-Aligned)	0.5	21.77 (25.00)	22.62 (23.38)	25.6 (31.6)	27.3 (30.0)
	1.0	21.85 (25.22)	22.78 (23.67)	25.3 (31.2)	27.0 (30.1)
	2.0	21.01 (24.88)	22.29 (23.80)	23.5 (30.9)	25.8 (30.4)
R2L (Left-Aligned)	0.5	23.75 (25.04)	23.15 (23.25)	27.6 (31.4)	27.8 (28.6)
	1.0	23.72 (25.29)	22.89 (22.89)	28.0 (31.6)	28.0 (29.3)
	2.0	24.09 (25.64)	23.61 (23.85)	28.6 (31.9)	28.3 (29.9)
R2L (Right-Aligned)	0.5	19.23 (23.52)	19.70 (21.02)	21.3 (31.3)	22.3 (28.3)
	1.0	19.56 (23.27)	20.20 (21.55)	20.9 (30.5)	21.6 (28.3)
	2.0	20.19 (23.55)	20.84 (22.22)	20.3 (30.9)	21.5 (28.7)
Common First	0.5	25.20 (25.43)	25.05 (25.05)	29.9 (31.2)	30.5 (30.5)
	1.0	25.46 (25.84)	25.76 (25.81)	30.5 (32.0)	31.1 (31.3)
	2.0	25.30 (25.76)	25.75 (25.83)	30.4 (32.2)	31.4 (31.9)
Rare First	0.5	22.83 (24.30)	23.19 (23.62)	27.0 (29.5)	28.7 (29.7)
	1.0	22.75 (24.56)	23.42 (23.99)	27.9 (30.7)	29.5 (30.5)
	2.0	23.10 (24.79)	24.00 (24.36)	28.1 (31.2)	29.8 (31.1)
Shortest First	0.5	24.93 (25.55)	24.94 (25.01)	27.4 (30.3)	29.1 (30.0)
	1.0	24.95 (25.72)	25.17 (25.28)	28.0 (30.9)	29.6 (30.8)
	2.0	25.05 (25.85)	25.26 (25.48)	28.2 (31.4)	30.3 (31.5)
Longest First	0.5	23.59 (25.09)	24.24 (24.56)	29.2 (31.4)	30.5 (31.2)
	1.0	23.53 (25.07)	24.68 (25.13)	29.2 (31.5)	31.0 (31.8)
	2.0	24.09 (25.78)	24.93 (25.37)	29.0 (31.9)	31.1 (32.1)
Alphabetical (A → Z → a → z)	0.5	24.49 (25.15)	24.87 (24.91)	29.2 (31.0)	30.1 (30.6)
	1.0	24.61 (25.19)	24.96 (25.12)	30.1 (32.0)	30.8 (31.4)
	2.0	24.77 (25.67)	25.45 (25.71)	29.7 (32.1)	30.7 (31.8)
Alphabetical (z → a → Z → A)	0.5	24.16 (25.24)	24.56 (24.73)	29.2 (31.4)	30.3 (30.8)
	1.0	24.19 (25.45)	24.65 (25.10)	29.3 (31.9)	30.7 (31.5)
	2.0	24.26 (25.76)	25.02 (25.40)	29.7 (32.3)	31.0 (32.0)
Easy First	0.5	22.58 (24.09)	22.16 (22.63)	27.5 (30.2)	28.4 (29.7)
	1.0	23.68 (25.08)	23.66 (24.03)	28.9 (31.6)	29.3 (30.7)
	2.0	23.87 (25.43)	24.64 (25.26)	29.1 (31.9)	30.4 (31.7)
Hard First	0.5	20.01 (23.46)	23.16 (23.61)	24.7 (29.7)	28.7 (30.2)
	1.0	20.96 (24.36)	23.76 (24.56)	25.4 (30.1)	29.1 (30.7)
	2.0	21.97 (24.90)	24.33 (24.70)	26.4 (31.1)	29.9 (31.4)

Table 4: Development BLEU results for WMT14 En-De newstest2013 and WMT18 En-Zh newstest2017. The first number in each column is the result obtained without an EOS penalty, while the second number in parentheses is the score obtained with the best EOS penalty for that setting.