

	fr-en (6268 pairs)	de-en (6382 pairs)	es-en (4106 pairs)	cz-en (2251 pairs)	hu-en (2193 pairs)	xx-en (1952 pairs)	Overall (23152 pairs)
ulc	.55	.56	.51	.50	.51	.51	.54
rte (absolute)	.54	.56	.51	.50	.55	.51	.53
wpF	.54	.55	.50	.47	.48	.51	.52
wcd6p4er	.54	.54	.49	.48	.48	.50	.52
maxsim	.53	.55	.49	.47	.50	.49	.52
bleusp	.54	.55	.49	.47	.46	.50	.51
bleusp4114	.53	.55	.48	.47	.46	.50	.51
rte (pairwise)	.49	.48	.52	.53	.55	.52	.51
terp	.52	.53	.48	.46	.45	.48	.50
meteor-0.6	.50	.53	.46	.48	.47	.47	.49
meteor-rank	.50	.52	.46	.48	.47	.47	.49
meteor-0.7	.49	.52	.46	.48	.47	.47	.49
ter	.48	.47	.43	.41	.40	.42	.45
wpbleu	.46	.45	.46	.39	.35	.45	.44

Table 10: Sentence-level consistency of the automatic metrics with human judgments for translations into English. Italicized numbers fall below the random-choice baseline.