

	en-fr (2967 pairs)	en-de (6563 pairs)	en-es (3249 pairs)	en-cz (11242 pairs)	Overall (24021 pairs)
Oracle	.62	.59	.63	.60	.60
terp	.62	.50	.59	.53	.54
ter	.61	.51	.58	.50	.53
bleusp	.62	.48	.59	.50	.52
bleusp4114	.63	.48	.59	.50	.52
wcd6p4er	.62	.46	.58	.50	.52
wpbleu	.63	.51	.60	n/a	.56
wpF	.63	.50	.59	n/a	.55

Table 13: Consistency of the automatic metrics when their system-level ranks are treated as sentence-level scores. Oracle shows the consistency of using the system-level human ranks that are given in Table 6.