The accuracyRank in each reference from each language pair in each domain (used for the universal metrics evaluation) is displayed in Tables B.1-B.12

	AccuracyRank	
	Reference A	Reference B
COMET	0.75	0.758
BLEU	0.733	0.733
chrF	0.758	0.75
ZeroEdit	0.625	0.742
TER	0.767	0.758
GLEU	0.825	0.825
BERTScore	0.825	0.8
Average	0.8	0.8
Median	0.758	0.75
Social Choice Theory	0.833	0.808
Weighted Mean 1000	0.817	0.8
Weighted Mean 3000	0.8	0.842
Weighted Mean 5000	0.792	0.817
Weighted Sum 1000	0.65	0.7
Weighted Sum 3000	0.575	0.558
Weighted Sum 5000	0.633	0.708

Table B.1: AccuracyRank from the language pair en-de in domain conversation.

	AccuracyRank	
	Reference A	Reference B
COMET	0.8	0.617
BLEU	0.692	0.558
chrF	0.65	0.533
ZeroEdit	0.667	0.55
TER	0.683	0.567
GLEU	0.658	0.5
BERTScore	0.675	0.567
Average	0.667	0.558
Median	0.683	0.567
Social Choice Theory	0.683	0.575
Weighted Mean 1000	0.683	0.55
Weighted Mean 3000	0.683	0.55
Weighted Mean 5000	0.733	0.583
Weighted Sum 1000	0.85	0.667
Weighted Sum 3000	0.55	0.483
Weighted Sum 5000	0.725	0.6

Table B.2: AccuracyRank from the language pair en-de in domain ecommerce.

	AccuracyRank	
	Reference A	Reference B
COMET	0.825	0.75
BLEU	0.675	0.658
chrF	0.583	0.567
ZeroEdit	0.65	0.558
TER	0.658	0.583
GLEU	0.667	0.608
BERTScore	0.667	0.65
Average	0.675	0.6
Median	0.658	0.583
Social Choice Theory	0.675	0.625
Weighted Mean 1000	0.692	0.608
Weighted Mean 3000	0.692	0.625
Weighted Mean 5000	0.725	0.667
Weighted Sum 1000	0.775	0.725
Weighted Sum 3000	0.642	0.542
Weighted Sum 5000	0.7	0.592

Table B.3: AccuracyRank from the language pair en-de in domain news.

	AccuracyRank	
	Reference A	Reference B
COMET	0.817	0.767
BLEU	0.742	0.667
chrF	0.725	0.608
ZeroEdit	0.733	0.592
TER	0.742	0.642
GLEU	0.758	0.717
BERTScore	0.767	0.692
Average	0.775	0.65
Median	0.742	0.642
Social Choice Theory	0.767	0.667
Weighted Mean 1000	0.758	0.667
Weighted Mean 3000	0.792	0.675
Weighted Mean 5000	0.775	0.717
Weighted Sum 1000	0.908	0.817
Weighted Sum 3000	0.642	0.508
Weighted Sum 5000	0.825	0.683

Table B.4: AccuracyRank from the language pair en-de in domain social.

	AccuracyRank
	Reference A
COMET	0.783
BLEU	0.767
chrF	0.808
ZeroEdit	0.725
TER	0.792
GLEU	0.758
BERTScore	0.783
Average	0.8
Median	0.792
Social Choice Theory	0.8
Weighted Mean 1000	0.808
Weighted Mean 3000	0.8
Weighted Mean 5000	0.792
Weighted Sum 1000	0.75
Weighted Sum 3000	0.483
Weighted Sum 5000	0.7

Table B.5: AccuracyRank from the language pair en-ru in domain conversation.

	AccuracyRank
	Reference A
COMET	0.783
BLEU	0.758
chrF	0.725
ZeroEdit	0.717
TER	0.7
GLEU	0.7
BERTScore	0.733
Average	0.75
Median	0.7
Social Choice Theory	0.75
Weighted Mean 1000	0.758
Weighted Mean 3000	0.775
Weighted Mean 5000	0.783
Weighted Sum 1000	0.825
Weighted Sum 3000	0.592
Weighted Sum 5000	0.792

Table B.6: AccuracyRank from the language pair en-ru in domain ecommerce.

	AccuracyRank
	Reference A
COMET	0.792
BLEU	0.85
chrF	0.708
ZeroEdit	0.758
TER	0.792
GLEU	0.733
BERTScore	0.85
Average	0.842
Median	0.792
Social Choice Theory	0.833
Weighted Mean 1000	0.817
Weighted Mean 3000	0.817
Weighted Mean 5000	0.875
Weighted Sum 1000	0.75
Weighted Sum 3000	0.625
Weighted Sum 5000	0.658

Table B.7: AccuracyRank from the language pair en-ru in domain news.

	AccuracyRank
	Reference A
COMET	0.917
BLEU	0.875
chrF	0.842
ZeroEdit	0.842
TER	0.85
GLEU	0.858
BERTScore	0.842
Average	0.858
Median	0.85
Social Choice Theory	0.85
Weighted Mean 1000	0.892
Weighted Mean 3000	0.858
Weighted Mean 5000	0.875
Weighted Sum 1000	0.875
Weighted Sum 3000	0.742
Weighted Sum 5000	0.942

Table B.8: AccuracyRank from the language pair en-ru in domain social.

	AccuracyRank	
	Reference A	Reference B
COMET	0.7	0.7
BLEU	0.75	0.717
chrF	0.7	0.742
ZeroEdit	0.717	0.783
TER	0.758	0.742
GLEU	0.658	0.6
BERTScore	0.725	0.742
Average	0.75	0.742
Median	0.758	0.742
Social Choice Theory	0.758	0.742
Weighted Mean 1000	0.75	0.742
Weighted Mean 3000	0.75	0.758
Weighted Mean 5000	0.733	0.733
Weighted Sum 1000	0.667	0.667
Weighted Sum 3000	0.583	0.542
Weighted Sum 5000	0.617	0.575

Table B.9: AccuracyRank from the language pair zh-en in domain conversation.

	AccuracyRank	
	Reference A	Reference B
COMET	0.892	0.908
BLEU	0.867	0.867
chrF	0.708	0.725
ZeroEdit	0.683	0.733
TER	0.792	0.758
GLEU	0.7	0.758
BERTScore	0.775	0.742
Average	0.775	0.775
Median	0.792	0.758
Social Choice Theory	0.775	0.767
Weighted Mean 1000	0.792	0.783
Weighted Mean 3000	0.817	0.8
Weighted Mean 5000	0.867	0.867
Weighted Sum 1000	0.892	0.908
Weighted Sum 3000	0.575	0.525
Weighted Sum 5000	0.842	0.883

Table B.10: AccuracyRank from the language pair zh-en in domain ecommerce.

	AccuracyRank	
	Reference A	Reference B
COMET	0.725	0.717
BLEU	0.75	0.717
chrF	0.717	0.733
ZeroEdit	0.692	0.7
TER	0.675	0.65
GLEU	0.708	0.592
BERTScore	0.658	0.65
Average	0.692	0.708
Median	0.675	0.65
Social Choice Theory	0.683	0.658
Weighted Mean 1000	0.708	0.708
Weighted Mean 3000	0.7	0.692
Weighted Mean 5000	0.692	0.692
Weighted Sum 1000	0.642	0.633
Weighted Sum 3000	0.717	0.708
Weighted Sum 5000	0.717	0.7

Table B.11: AccuracyRank from the language pair zh-en in domain news.

	AccuracyRank	
	Reference A	Reference B
COMET	0.658	0.733
BLEU	0.642	0.708
chrF	0.592	0.717
ZeroEdit	0.6	0.65
TER	0.6	0.667
GLEU	0.683	0.633
BERTScore	0.583	0.65
Average	0.608	0.683
Median	0.6	0.667
Social Choice Theory	0.608	0.675
Weighted Mean 1000	0.608	0.708
Weighted Mean 3000	0.617	0.7
Weighted Mean 5000	0.633	0.7
Weighted Sum 1000	0.733	0.692
Weighted Sum 3000	0.592	0.55
Weighted Sum 5000	0.692	0.783

Table B.12: AccuracyRank from the language pair zh-en in domain social.