

Table A1. Statistical Test across Six Translation Datasets for RQ2.

	Computational Accuracy (CA)		Exact Match Accuracy (EM Acc)	
UniTrans (LLaMA-7B)	LLaMA-7B		LLaMA-7B	
	p-value	Cliff's Delta	p-value	Cliff's Delta
	0.0139	0.444 - Large	0.0139	0.444 - Large
UniTrans (LLaMA-13B)	LLaMA-13B		LLaMA-13B	
	p-value	Cliff's Delta	p-value	Cliff's Delta
	0.0139	0.889 - Large	0.0139	0.389 - medium
UniTrans (GPT-3.5)	GPT-3.5		GPT-3.5	
	p-value	Cliff's Delta	p-value	Cliff's Delta
	0.0139	0.278 - small	0.0139	0.306 - medium

Table A2. Statistical Test across Six Translation Datasets for RQ3.

	Computational Accuracy (CA)				Exact Match Accuracy (EM Acc)	
	LLaMA-7B		+TAP		LLaMA-7B	
	p-value	Cliff's Delta	p-value	Cliff's Delta	p-value	Cliff's Delta
+TAP	0.0139	0.389 - medium	/	/	0.0139	0.417 - medium
+TRP	/	/	0.0139	0.222 - small	/	/
	LLaMA-13B		+TAP		LLaMA-13B	
	p-value	Cliff's Delta	p-value	Cliff's Delta	p-value	Cliff's Delta
+TAP	0.0139	0.778 - Large	/	/	0.0139	0.333 - medium
+TRP	/	/	0.0139	0.278 - small	/	/
	GPT-3.5		+TAP		GPT-3.5	
	p-value	Cliff's Delta	p-value	Cliff's Delta	p-value	Cliff's Delta
+TAP	0.0370	0.149 - small	/	/	0.0370	0.278 - small
+TRP	/	/	0.0139	0.194 -small	/	/