

Figure 9: F1 score results of supervised fine-tuning and few-shot prompting on the test set of README++.

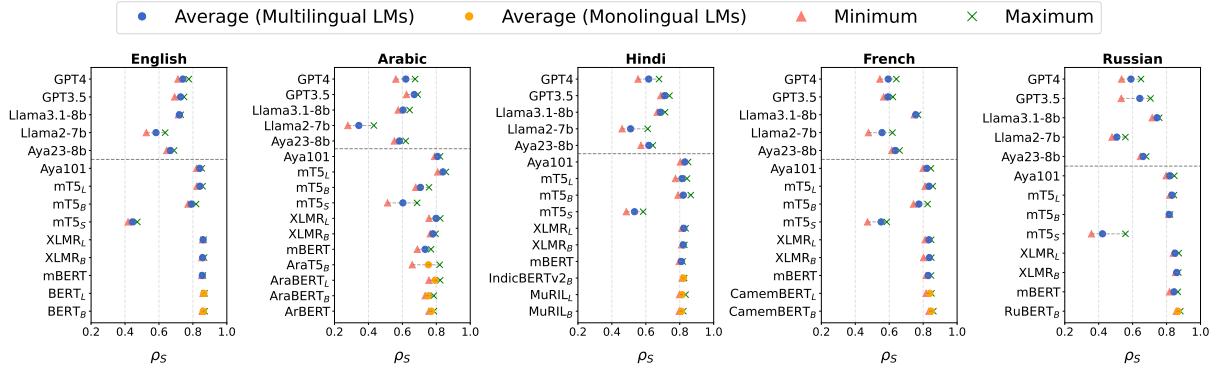


Figure 10: Spearman Correlation (ρ_s) of supervised fine-tuning and few-shot prompting on the test set of README++.

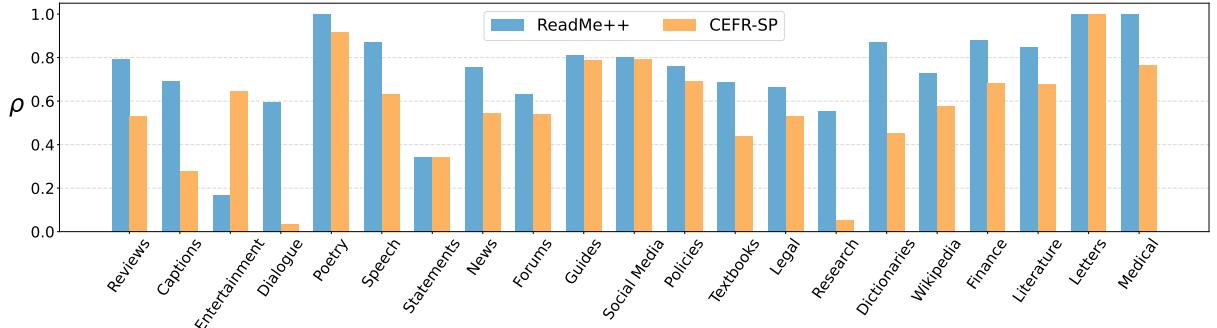


Figure 11: Pearson Correlation per domain for XLMR_L trained using README++ and CEFR-SP. The model trained with README++ achieves better domain generalization, shown by higher correlation in all but one domain (Entertainment).

F Annotation Interface

Figures 16 and 17 show screenshots of our developed annotation interface for English sentences, where annotators perform a rank-and-rate approach to assign readability scores to 5 sentences in each batch. Annotators are asked to first rank sentences which they can do by simply dragging them. They are then asked to choose a rating for each sentence from a drop-down list. For each sentence, we provide the option to show its context, which shows

the sentence in the paragraph to which it belongs. Figures 18 and 19 show screenshots of the interface for Arabic and Hindi respectively. An additional button to mark transliterations is added.

G License and Use Terms

We provide in Tables 18, 19, and 20 the license or usage term for each data source used in the creation of the corpus as follows:

- License: exact license under which data is avail-

Model	ReadMe++		CEFR-SP		CompDS	
	F1	ρ	F1	ρ	F1	ρ
en → ar						
mBERT	19.94	0.512	12.38	0.368	1.76	0.099
XLM-R _B	32.63	0.645	9.61	0.068	7.21	0.120
XLM-R _L	31.48	0.606	8.81	0.071	5.99	0.322
en → hi						
mBERT	15.13	0.521	8.72	0.375	6.45	0.171
XLM-R _B	16.57	0.655	9.87	0.146	9.81	0.398
XLM-R _L	23.87	0.702	13.15	0.267	10.38	0.381
en → fr						
mBERT	30.63	0.751	10.87	0.490	8.02	0.341
XLM-R _B	33.96	0.746	10.37	0.091	8.97	0.399
XLM-R _L	30.29	0.768	11.06	-0.026	5.92	0.335
en → ru						
mBERT	16.25	0.610	9.11	0.479	10.9	0.396
XLM-R _B	21.27	0.671	13.16	0.253	12.64	0.404
XLM-R _L	24.60	0.760	15.69	0.173	10.33	0.412
en → it						
mBERT	12.79	0.270	7.91	0.248	10.37	0.119
XLM-R _B	14.38	0.295	9.66	0.029	12.00	0.137
XLM-R _L	14.68	0.239	9.88	-0.043	10.06	0.099
en → de						
mBERT	15.98	0.672	12.51	0.595	6.88	0.347
XLM-R _B	27.13	0.702	14.02	0.196	8.68	0.529
XLM-R _L	22.19	0.701	10.00	-0.092	11.84	0.408

Table 11: Zero-shot cross-lingual transfer performance. Models fine-tuned on English data (en) of README++ significantly outperform models fine-tuned with CEFR-SP (Arase et al., 2022) or CompDS (Brunato et al., 2018) for Arabic (ar), Hindi (hi), Italian (it), and German (de).

able (CC BY 4.0 or other).

- Public Domain: data available in the public domain.
- Personal/Non-Commercial: source grants usage permission of data for personal/non-commercial purposes.
- (—): denotes that data needs to be requested from authors.

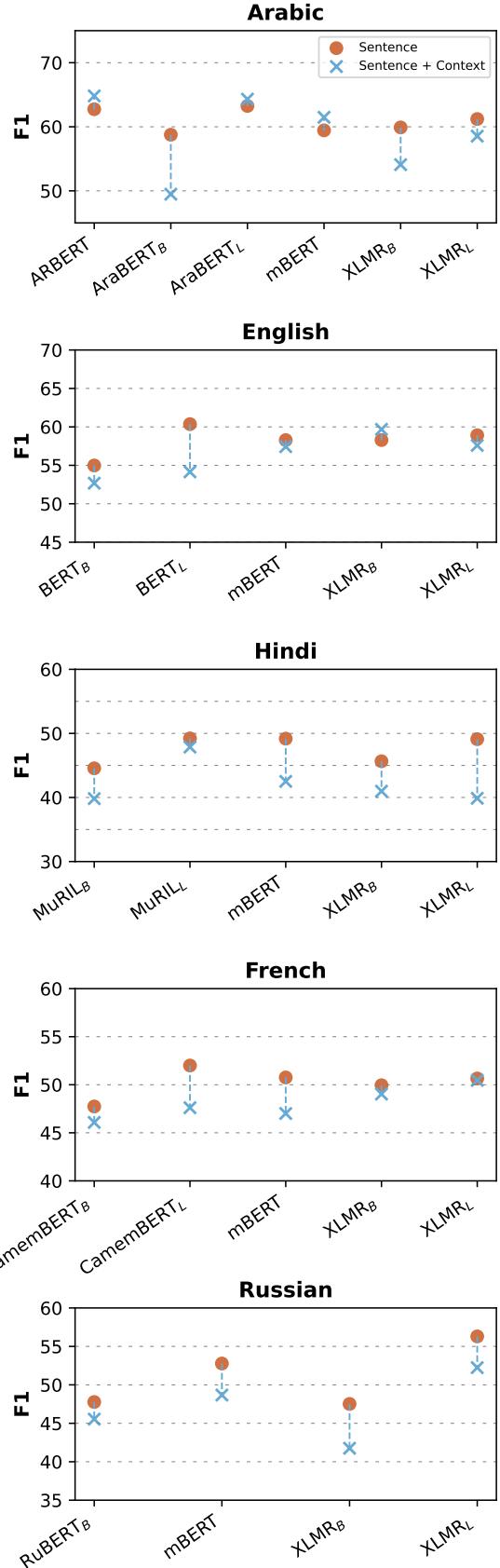


Figure 12: Effect of providing context during fine-tuning.

Domain	en → ar		en → fr		en → hi		en → ru	
	ReadMe++	CEFR-SP	ReadMe++	CEFR-SP	ReadMe++	CEFR-SP	ReadMe++	CEFR-SP
	0.545	0.165	0.551	0.179	0.336	0.028	0.644	0.202
Captions	0.126	0.269	0.635	-0.387	0.438	0.122	0.150	-0.220
Dialogue	-0.274	0.000	—	—	—	—	—	—
Dictionaries	0.374	0.107	0.000	0.000	0.657	0.099	0.397	0.288
Entertainment	—	—	0.784	-0.013	—	—	0.352	-0.084
Finance	0.440	0.161	0.564	0.000	0.603	0.281	0.737	-0.109
Forums	0.534	0.024	0.388	-0.030	0.362	0.041	0.438	0.011
Guides	0.277	-0.093	0.557	-0.190	0.362	0.261	0.782	-0.220
Legal	—	—	0.794	0.000	—	—	0.892	0.214
Letters	0.692	0.081	0.709	-0.368	0.561	0.168	0.498	0.059
Literature	0.447	0.000	—	—	—	—	—	—
News	0.000	0.000	0.339	-0.068	0.202	-0.347	0.779	0.112
Poetry	0.835	0.009	0.727	-0.070	0.551	-0.427	0.703	0.144
Policies	0.562	-0.021	0.564	0.154	0.501	-0.112	0.647	0.262
Research	0.620	0.313	0.489	-0.677	0.341	0.036	0.452	-0.106
Social Media	0.337	-0.147	0.618	0.291	0.668	0.200	0.583	0.118
Speech	0.374	-0.019	0.592	-0.193	0.331	-0.013	0.602	-0.130
Statements	0.600	0.569	—	—	0.427	-0.201	—	—
Textbooks	0.570	0.240	—	—	0.375	-0.018	0.000	-0.196
User Reviews	0.644	0.111	0.625	0.097	0.630	0.110	0.715	0.109

Table 12: Pearson Correlation per domain when performing cross lingual transfer to Arabic, French, Hindi, and Russian using XLMR_L fine-tuned with README++ (en) vs CEFR-SP-WikiAuto (Arase et al., 2022).