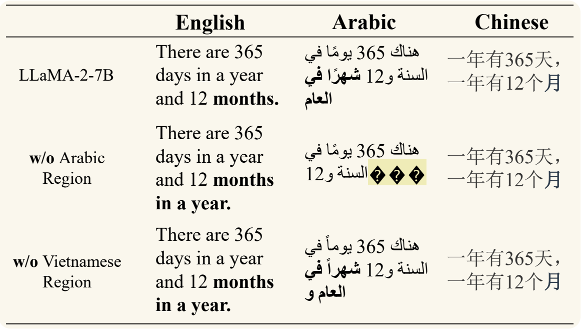

	English	Arabic	Chinese
LLaMA-2-7B	There are 365 days in a year and 12 **months.**	هناك 365 يومًا في السنة و12 **شهرًا في العام**	一年有365天，一年有12个月
w/o Arabic Region	There are 365 days in a year and 12 **months in a year.**	هناك 365 يومًا في السنة و12���	一年有365天，一年有12个月
w/o Vietnamese Region	There are 365 days in a year and 12 **months in a year.**	هناك 365 يوماً في السنة و12 **شهراً في العام و**	一年有365天，一年有12个月

Figure 8: Model's generation with monolingual regions removed. Here, we use "*There are 365 days in a year and 12*" as prompt input, and translate it into Arabic and Chinese to evaluate model's performance in three languages.

Languages		Arabic (**10K**)		Arabic (**100K**)	
	Base	Top	Bottom	Top	Bottom
Arabic	6.771	**81.659**	6.785	**135.02**	6.786
Chinese	8.562	9.309	8.593	9.165	8.588
Italian	14.859	16.61	14.959	16.366	14.919
Japanese	10.888	12.238	10.932	11.956	10.923
Korean	4.965	5.534	4.972	5.442	4.969
Persian	6.509	**34.142**	6.52	**43.414**	6.508
Portuguese	15.318	16.909	15.262	16.86	15.239
Russian	12.062	13.708	12.145	13.781	12.141
Spanish	17.079	18.543	17.24	18.314	17.2
Ukrainian	9.409	11.243	9.433	11.225	9.439
Vietnamese	5.824	6.412	5.874	6.335	5.871

Table 10: LLaMA-2-7B perplexity on 11 languages with an 'Arabic' region removal. Here, 'Arabic' and 'Persian' are gray-filled while others are unfilled, 'Top' and 'Bottom' are deduplicated, and 'Base' is unchanged. Values with greater changes compared to the other regions' removals are in bold.

Languages		Spanish (**10K**)		Spanish (**100K**)	
	Base	Top	Bottom	Top	Bottom
Arabic	6.771	7.158	6.788	7.15	6.789
Chinese	8.562	8.984	8.594	8.971	8.596
Italian	14.859	**21.292**	14.933	**27.004**	14.95
Japanese	10.888	11.376	10.913	11.426	10.933
Korean	4.965	5.169	4.967	5.167	4.972
Persian	6.509	6.906	6.484	6.945	6.529
Portuguese	15.318	**21.217**	15.249	**26.877**	15.256
Russian	12.062	13.039	12.133	13.252	12.141
Spanish	17.079	**38.876**	17.224	**64.513**	17.225
Ukrainian	9.409	10.027	9.439	10.082	9.439
Vietnamese	5.824	6.136	5.875	6.145	5.877

Table 11: LLaMA-2-7B perplexity on 11 languages with a 'Spanish' region removal. Here, 'Spanish', 'Italian' and 'Portuguese' are gray-filled while others are unfilled, and values with greater changes compared to the other regions' removals are in bold.

Languages		Chinese (**10K**)		Chinese (**100K**)	
	Base	Top	Bottom	Top	Bottom
Arabic	6.771	7.161	6.79	7.714	6.784
Chinese	8.562	**10.899**	8.592	**12.079**	8.586
Italian	14.859	16.041	14.939	15.881	14.932
Japanese	10.888	**12.265**	10.922	**12.878**	10.904
Korean	4.965	5.343	4.974	5.341	4.960
Persian	6.509	6.92	6.519	6.865	6.516
Portuguese	15.318	16.285	15.27	16.241	15.26
Russian	12.062	12.887	12.136	12.973	12.145
Spanish	17.079	18.068	17.216	17.974	17.219
Ukrainian	9.409	10.144	9.439	10.207	9.447
Vietnamese	5.824	6.261	5.878	6.296	5.870

Table 12: LLaMA-2-7B perplexity on 11 languages with a 'Chinese' region removal. Here, 'Chinese' and 'Japanese' are gray-filled while others are unfilled, and values with greater changes compared to the other regions' removals are in bold.

Languages		Korean (**10K**)		Korean (**100K**)	
	Base	Top	Bottom	Top	Bottom
Arabic	6.771	7.259	6.791	7.316	6.783
Chinese	8.562	9.14	8.594	9.173	8.594
Italian	14.859	15.91	14.941	15.791	14.938
Japanese	10.888	**13.273**	10.919	**15.062**	10.932
Korean	4.965	**8.364**	4.971	**13.128**	4.971
Persian	6.509	7.38	6.522	7.574	6.522
Portuguese	15.318	16.113	15.259	15.984	15.26
Russian	12.062	12.758	12.138	12.827	12.136
Spanish	17.079	17.981	17.214	17.858	17.225
Ukrainian	9.409	10.065	9.434	10.108	9.442
Vietnamese	5.824	6.188	5.874	6.177	5.874

Table 13: LLaMA-2-7B perplexity on 11 languages with a 'Korean' region removal. Here, 'Korean' and 'Japanese' are gray-filled while others are unfilled, and values with greater changes compared to the other regions' removals are in bold.

Languages		Vietnamese (**10K**)		Vietnamese (**100K**)	
	Base	Top	Bottom	Top	Bottom
Arabic	6.771	7.435	6.785	7.341	6.789
Chinese	8.562	9.576	8.589	9.372	8.592
Italian	14.859	16.979	14.952	16.497	14.937
Japanese	10.888	12.027	10.946	11.814	10.941
Korean	4.965	5.44	4.97	5.335	4.979
Persian	6.509	7.315	6.501	7.243	6.521
Portuguese	15.318	17.159	15.249	16.805	15.258
Russian	12.062	13.107	12.141	13.007	12.144
Spanish	17.079	18.801	17.244	18.369	17.233
Ukrainian	9.409	10.316	9.447	10.217	9.433
Vietnamese	5.824	**24.382**	5.872	**27.817**	5.874

Table 14: LLaMA-2-7B perplexity on 11 languages with a 'Vietnamese' region removal. Here, 'Vietnamese' is gray-filled while others are unfilled, and values with greater changes compared to the other regions' removals are in bold.

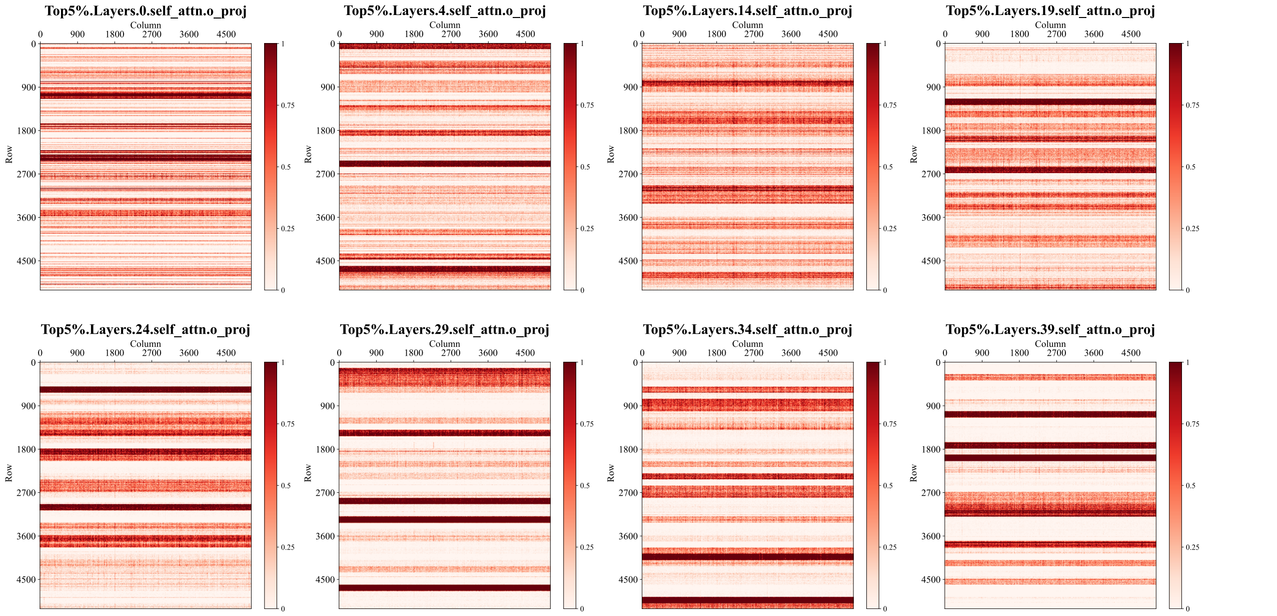

Figure 9: Visualization of the linguistic competence region (the 'Top' 5% region) in Attention.o matrix across 8 different layers. The scale from 0 to 1 (after normalization) represent the proportion of parameters within a 3 × 3 vicinity that belong to the 'Top' region.

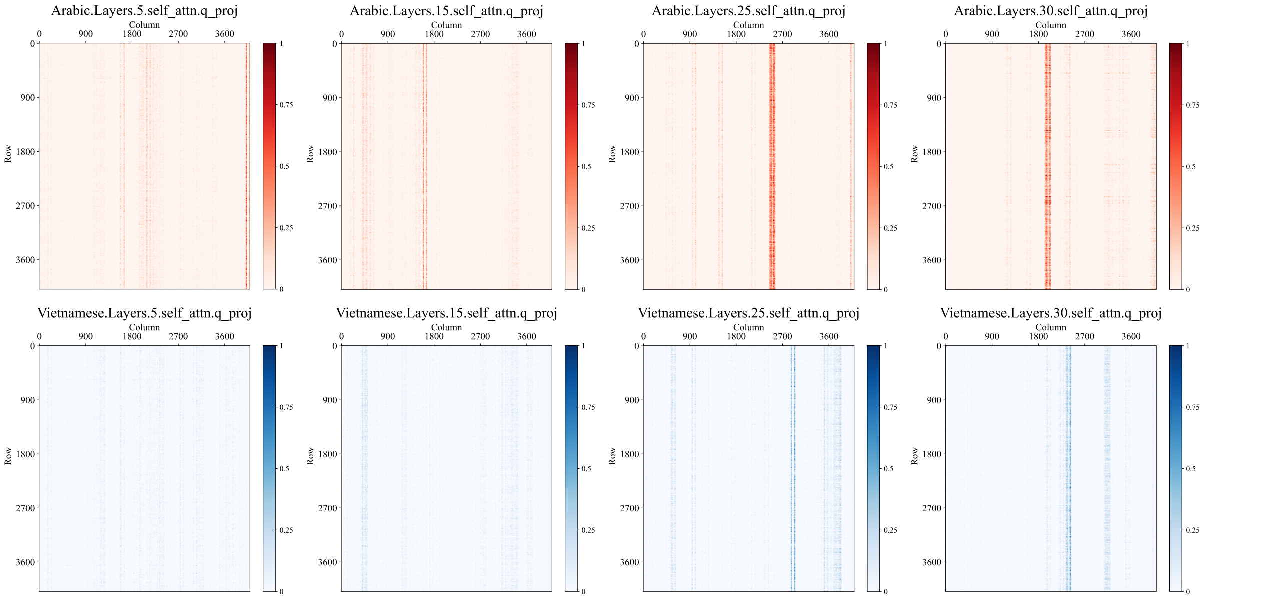

Figure 10: Visualization of the monolingual regions for 'Arabic' and 'Vietnamese' across 4 different layers in the Attention.q matrix. The scale from 0 to 1 (after normalization) represent the proportion of parameters within a 3 × 3 vicinity that belong to the monolingual regions.

Languages	LLaMA-2-7B **3% (100K)**				LLaMA-2-13B **3% (100K)**			
	Base	Top	Bottom	Random	Base	Top	Bottom	Random
Arabic	6.771	127208.250	6.772	7.895	6.261	102254.758	6.316	7.112
Chinese	8.652	295355.5	8.565	9.837	7.838	84086.906	7.806	8.619
Czech	19.834	62692.367	19.835	24.005	17.744	56102.227	17.650	20.485
Danish	8.372	47654.156	8.372	9.929	7.402	47213.586	7.401	8.278
Dutch	16.959	48478.594	16.959	20.121	15.64	46303.559	15.572	18.295
English	7.653	16573.422	7.653	8.359	7.447	25212.217	7.234	7.821
Finnish	7.566	45711.992	7.566	8.934	6.887	48811.242	6.861	7.826
French	13.605	48268.211	13.605	15.003	12.765	45674.492	12.573	13.682
German	18.355	64015.117	18.356	15.404	17.29	51692.125	16.973	18.972
Greek	3.832	224595.781	3.833	4.527	3.599	80657.891	3.599	4.146
Hungarian	16.365	52828.691	16.363	20.039	14.756	58107.137	14.834	17.633
Indonesian	44.269	33121.945	44.318	48.175	37.909	51611.625	37.838	38.548
Italian	14.859	58908.879	14.860	17.341	13.694	47375.844	13.730	15.207
Japanese	10.888	322031.406	10.896	12.535	10.072	75236.031	10.137	11.661
Korean	4.965	125345.359	4.967	5.649	4.724	90768.844	4.743	5.241
Malay	66.581	22603.727	66.843	74.167	46.885	40468.750	46.912	58.947
Malayalam	5.133	373710.188	5.134	6.396	4.972	16990.266	4.972	5.654
Norwegian	14.425	31526.176	14.427	17.854	13.142	45820.109	13.139	15.041
Persian	6.509	81959.719	6.511	7.628	6.205	92201.812	6.229	7.009
Polish	12.629	66906.469	12.629	14.843	11.414	55923.156	11.311	12.987
Portuguese	15.318	47763.059	15.319	17.297	13.667	51498.402	13.982	15.376
Romanian	10.893	43498.008	10.895	13.061	9.652	54986.055	9.693	10.969
Russian	12.062	170776.750	12.064	13.728	11.048	112574.609	10.948	11.757
Spanish	17.079	51940.859	17.082	18.98	16.351	54005.891	16.138	17.292
Swahili	75.908	29234.168	75.892	89.380	70.519	48802.227	70.402	81.216
Swedish	14.714	49425.969	14.714	17.258	13.229	48622.266	13.337	14.933
Tamil	4.162	381070.844	4.162	5.04	4.028	111060.516	4.049	4.488
Turkish	11.214	46986.391	11.215	13.765	9.834	50303.562	9.763	11.374
Ukrainian	9.409	120719.938	9.409	10.875	8.295	116287.305	8.297	9.076
Vietnamese	5.824	40126.527	5.824	6.614	5.471	42336.426	5.437	5.995

Table 15: LLaMA-2 perplexity on 30 languages with 3% removal ratio. '100K' means that the region is selected from 100,000 samples. 'Top' and 'Bottom' respectively indicate the N parameters with the highest and lowest cumulative $\mathcal{I}_j^*(\theta)$ during the further pre-training across the six languages. 'Random' denotes the randomly selecting N while 'Base' represents no removal. Here, N equals 3% of the total number in each parameter matrix.