Lang	Tokens (M)
English	803,527
French	77,420
Japanese	66,054
Romanian	24,176
Finnish	16,804
Thai	10,842
Basque	105

Table 4: The statistics of language resource level for 7 languages used in this work.

- 653 out of 4397 for French (roughly 15%)
- 449 out of 4397 for Romanian (roughly 10%)
- 243 out of 4397 for Basque (roughly 5%)
- 172 out of 4397 for Finnish (roughly 4%)

This ratio remains similar when counted only on the test split. We can observe from above that the ratio is quite limited.

B Full Experimental Results

In the Appendix, we report our full experimental results across different models with varying model sizes, seed dictionary sizes, different k-values for P@K, in following figure and tables (Figure 3 and Table 4-23). These results provide full scope of our analysis, allowing for an in-depth comparison of model performances.

For different models, we use their HuggingFace PyTorch implementation[10]. For Procrustes Analysis, we utilize the MUSE[11] package. All experiments are run on a single NVIDIA A100 GPU.

[10] https://huggingface.co/bigscience/bloomz-{1b7,3b,7b1}, https://huggingface.co/meta-llama/Llama-2-{7,13,70}b-chat-hf, https://huggingface.co/bigscience/mt0-{large,xl,xxl}, https://huggingface.co/CohereForAI/aya-101

[11] https://github.com/facebookresearch/MUSE