**Language and database**

*Nav*

Korean Romanization <https://languagelog.ldc.upenn.edu/nll/?p=51316>

Korean lexical database <https://www.cjk.org/data/korean/nlp/korean-lexical-database/>

Korean translation <https://www.kaggle.com/datasets/rareloto/naver-dictionary-conversation-of-the-day>

Chinese Romanization <https://en.wikipedia.org/wiki/Pinyin>

Chinese translation <https://www.kaggle.com/datasets/concyclics/machine-translation-chinese-and-english?select=chinese.zh>

Multiple languages <https://www.kaggle.com/datasets/rtatman/sentiment-lexicons-for-81-languages>

Japanese and Hebrew <https://www.kaggle.com/datasets/dhruvildave/wikibooks-dataset>

<https://github.com/AI4Bharat/IndicXlit/tree/master>

LLAMA2: <https://github.com/meta-llama/llama>

IndiXlit: <https://github.com/AI4Bharat/IndicXlit>

<https://github.com/allenai/open-instruct/tree/main>

Paul

Japanese: <https://paperswithcode.com/dataset/jesc>

Anusha

Llama

<https://replicate.com/blog/fine-tune-translation-model-axolotl>

<https://medium.aiplanet.com/no-code-llm-fine-tuning-using-axolotl-2db34e3d0647>

<https://medium.com/@kevaldekivadiya2415/fine-tuning-llama-2-your-path-to-chemistry-text-perfection-aa4c54ff5790>

Transliteration of the Cyrillic script

<https://github.com/opendatakosovo/cyrillic-transliteration>