Sophistication

Step 1 Concatenate all corpus files into one txt file in the data folder

bash concatenate_files.sh ~/switchdrive/IMAGINE_files/chatGPT/project_2/final_files_simple_prompt

Step 2 Run the main script

bash sophistication.sh

Lexical richness

bash lxr_scores.sh ~/switchdrive/IMAGINE_files/chatGPT/project_2/final_files_simple_prompt/{corpus}

Morphology

Step 1 Extract vocabulary of most frequent words

bash create_most_freq_vocs.sh ~/switchdrive/IMAGINE_files/chatGPT/project_2/final_files_simple_prompt/{corpus}

Step 2 Run diversity analysis

for single file

python3 shannon_pairwise.py -f ~/switchdrive/IMAGINE_files/datasets/wmtnews21/wmtnews_test_de_A.txt -l de -sys A_wmt -v freq_voc/wmtnews_test_de_A.freq_voc > test.txt

for multiple files in a directory if considering the top 1000 most frequent lemmas with more than 1 morphological form

lang = {"en", "de"}

bash shannon_1000_mostfrequent_script.sh ~/switchdrive/IMAGINE_files/chatGPT/project_2/final_files_simple_prompt/{corpus} lang

if chosing all lemmas with more than one morphological form

lang = {"en", "de"}

bash mrph_all.sh ~/switchdrive/IMAGINE_files/chatGPT/project_2/final_files_simple_prompt/{corpus} lang