1.2.7 Normalization

Normalization is the task of "translating" non-standard language data into standard language. It can be performed manually or automatically with computational linguistics tools.

In the case of our corpus, we have manually normalized some data in the Swiss German dialect, resulting in the corpus WUS_DIALOG_GSW (5 chats, 34,683 tokens).

From:
https://whatsup.linguistik.uzh.ch/ -

Permanent link:
https://whatsup.linguistik.uzh.ch/01_corpus/02_preprocessing/07_normalization

Last update: **2022/06/27 07:21**