

Normalization

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

Another set of data was process automatically. You can read more about that project in:

Ruzsics, Tatiana; Luseti, Massimo; Göhring, Anne; Samardžić, Tanja; Stark, Elisabeth (2019): Neural Text Normalization with Adapted Decoding and PoS Features. [Natural Language Engineering](#).

This data will be made available soon.

From:

<https://whatsup.linguistik.uzh.ch/> -

Permanent link:

https://whatsup.linguistik.uzh.ch/01_corpus/02_preprocessing/07_normalization?rev=1587047796

Last update: **2022/06/27 09:21**

