1. THE CORPUS

The corpus consists of 617 chats that were sent in by the Swiss population in 2014 through a fixed procedure that was communicated in the press in order to get people interested. The individual chats were checked for their permission to use them and chats that did not have it were removed. Furthermore, available demographic data were linked to the chats.

Next processing steps comprised anonymization, the annotation of a main language per chat and thus the creation of subcorpora, application of further annotations (for languages, i.e. each message was annotated for its most likely language as opposed to the chat annotation performed in the first step), part of speech annotations, normalization for part of the dialectal Swiss German data.

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

Permanent link:
https://whatsup.linguistik.uzh.ch/01_corpus/start?rev=1586870848

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