

# 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.

In a first step the most basic processing of the data took place such as to allow the project members to work with the data. This included the [anonymization](#) and the annotation of a [main language](#) per chat and thus the creation of [subcorpora](#).

In a later step, more [annotations](#) were applied to the corpus. This included a more profound annotation of [languages](#) (i.e. each message was annotated for its language as opposed to the chat annotation performed in the first step), [part of speech annotations](#) were applied and the German dialectal data was [normalized](#).

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

Permanent link:  
[https://whatsup.linguistik.uzh.ch/01\\_corpus/start?rev=1586870548](https://whatsup.linguistik.uzh.ch/01_corpus/start?rev=1586870548)

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

