1.3.4 Languages and varieties

1.3.4 Languages and varieties

In order to assign a language tagging to each chat, student helpers read through the first 250 messages and assigned two possible attributes per language:

- lang_100_and_more
- lang_less_than_100

for the following languages:

- fra: French
- ita: Italian
- roh: Any variety of Romansh
- gsw: dialectal German as used in Switzerland
- deu: non-dialectal German
- eng: English
- spa: Spanish
- sla: Any Slavic language

Please note: In the browsing tool ANNIS, we created sub-corpora per language, where each message appears in one and only one sub-corpus, even though there may be several languages annotated as lang_100_and_more for a specific chat. If you want to work with all chats that contain a specific language in more than 100 messages, use the query *msg & meta::lang_100_and_more="fra, gsw"* on the whole corpus.

For an overview over languages in the corpus consult: Ueberwasser, Simone; Stark, Elisabeth (2017)2017: "What's up, Switzerland? A corpus-based research project in a multilingual country". In: Linguistik online, 84/5, 105-126. https://bop.unibe.ch/linguistik-online/article/view/3849/5834