"What's up, Switzerland?"

The project

The data underlying the corpus was collected in 2014 to constitute the data base of the research project "What's up, Switzerland?" under the lead of Prof. Elisabeth Stark (University of Zurich). The project was funded by the Swiss National Fund (Sinergia: CRSII1_160714) with CHF 1'832'647 and ran between 2016 - 2020. More about the project ...

Using the corpus

This corpus is freely available for academic, non-commercial research. When using the corpus, please make sure to quote correctly.

The corpus

Our authentic WhatsApp chats were gathered in summer 2014. Not all made it into the corpus (e.g. doublets, chats or message without permission etc.). In its present form, the corpus comprises:

- Number of chats: 617
- Number of messages (with permission to be used): 763’644
- Number of informants (who gave their permission): 944
- Number of tokens: 5'155'476 (without redactedQ.* (cf. Messages without permission))
- Number of emojis: 382'116

The corpus is built up of chats in all four national languages of Switzerland, i.e. Swiss German dialect, non-dialectal German, French, Italian and varieties of Romansh. In more detail, the following languages and varieties can be found in the corpus:

Available 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

Romansh varieties:

- roh-ja: Jauer Romansh
- roh-sr: Romontsch Sursilvan
- roh-st: Rumàntsch Sutsilvan

- roh-sm: Rumantsch Surmiran
- roh-pt: Rumauntsch Puter
- roh-vl: Rumantsch Vallader
- roh-gr: Rumantsch Grischun

More information about the corpus can be found in the section corpus and in the following publication:

Ueberwasser, Simone/Stark, Elisabeth (2017). "What's up, Switzerland? A corpus-based research project in a multilingual country". Linguistik online 84/5, 105-126 DOI: https://doi.org/10.13092/lo.84.3849 .

Quoting

When using the corpus, please quote as follows:

The corpus

Stark, Elisabeth; Ueberwasser, Simone; Göhring, Anne (2014-2020). *Corpus* "*What's up, Switzerland?*". University of Zurich. www.whatsup-switzerland.ch.

This documentation

Stark, Elisabeth; Ueberwasser, Simone (2020): The corpus "What's up, Switzerland?". Documentation, facts and figures. www.whatsup-switzerland.ch.

Creation of the corpus

Ueberwasser, Simone; Stark, Elisabeth (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

The project

Stark, Elisabeth (2016-2020). *SNSF project* "*What's up, Switzerland?*" (Sinergia: CRSII1_160714). University of Zurich. www.whatsup-switzerland.ch.

Raw data

If you want to use our raw data for computational linguistic projects, please contact Prof. Elisabeth Stark to see whether your project complies with our requirements. If we make the data available, a CC BY-NC-ND license is applied.

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

Permanent link:
https://whatsup.linguistik.uzh.ch/start?rev=1663003187

Last update: **2022/09/12 17:19**