

The Swiss SMS Corpus

The Swiss SMS corpus is one of the results of a project funded by the [Swiss National Science Foundation](#) between 2011 and 2014. Other results of the project are six dissertations as well as an abundance of student papers and publications. An overview over the whole project with can also be found in the [SNSF research database P3](#). The Swiss SMS corpus consists of 25'947 SMS (~650'000 tokens), which were sent in by the Swiss public in 2009/2010. Of all SMS, 41% are in Swiss German (dialect), 28% in non-dialectal German, 18% in French, 6% in Italian, and 4% in Romansh.

These data are freely [available](#) for bonafide academic research (CC-NY-NC), but not for commercial use. If you use the corpus, you agree to our conditions, i.e. to:

- Not use the data for commercial use, i.e. only for bonafide research
- Quote the source of the data as "Swiss SMS corpus" with the source as shown in the footer of this document and with a link to <https://sms.linguistik.uzh.ch>

This corpus would not be here without the following people/institutions, to whom we express our gratitude:

- The Swiss National Science Foundation financed the project and the dissertations over four years.
- The UZH [Zurich Center for Linguistics](#) supported us throughout the project with their IT knowledge and hosts the corpus.
- The UZH [Linguistic Research Infrastructure Project](#) hosts this project.
- [Anke Lüdeling](#) and her team allow us to use ANNIS as a browsing tool.

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

Permanent link:
<https://sms.linguistik.uzh.ch/start?rev=1641208933>

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

