# 2.4.5 Fields available

Below you find all fields that can be queried in the corpus in four categories depending on whether they relate to the chat, to the message, to the token or to the demographic meta data.

Please do not forget the Part of Speech annotations per language.

Hint: If you want to find a specific field, use the search function of your browser:

## Chat annotations

| **name** | example | |
|---|---|---|
| consent_speakers | node & meta::consent_speakers="2" | Messages in chats where two and exactly two people gave their permission for their messages to be used. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| contains_deu | node & meta::contains_deu="true" | Get all messages in chats that we have identified to contain non-dialectal German. |
| contains_eng | node & meta::contains_eng="true" | Get all messages in chats that we have identified to contain English. |
| contains_fra | node & meta::contains_fra="true" | Get all messages in chats that we have identified to contain French. |
| contains_gsw | node & meta::contains_gsw="true" | Get all messages in chats that we have identified to contain dialectal German. |
| contains_ita | node & meta::contains_ita="true" | Get all messages in chats that we have identified to contain Italian. |
| contains_roh | node & meta::contains_roh="true" | Get all messages in chats that we have identified to contain Romansh. |
| contains_sla | node & meta::contains_sla="true" | Get all messages in chats that we have identified to contain Slavic languages. |
| contains_spa | node & meta::contains_spa="true" | Get all messages in chats that we have identified to contain Spanish. |
| content_msg | node & meta::content_msg="818" | Find all messages in chats with exactly 818 messages for which we have the permission to use. This is an alphanumeric field, you cannot query for more or less than |
| demographics | node & meta::demographics="2" | Find all messages in chats where we have demographic data for exactly two participants. This is an alphanumeric field, we cannot query for more or less than |
| doc | node & meta::doc="chat126" | Find all messages in the chat 126 |

| | | |
|---|---|---|
| lang_100_and_more | node & meta::lang_100_and_more="deu, gsw" | Find all messages in chats with more than 100 messages in non-dialectal or dialectal German. The same query can be applied for fewer or more languages by separating them with commas as shown in the example. Other languages are fra and ita for French and Italian respectively as well as roh for Romansh. |
| lang_less_than_100 | node & meta::lang_less_than_100="roh" | Find all messages in chats with more than 100 messages in Romansh. The same query can be applied for fewer or more languages by separating them with commas as shown in the example. Other languages are fra and ita for French and Italian respectively as well as gsw for dialectal German as well as deu for non-dialectal German |
| no_consent_msg | node & meta::no_consent_msg="54" | Find all messages in chats with exactly 54 messages without consent to be used |
| speakers | node & meta::speakers="2" | Find all messages in chats with exactly two speakers regardless of whether we have their permission or not. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| total_msg | node & meta::total_msg="2443" | Find all messages in chats with exactly 2443 messages. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| user_msg | node & meta::user_msg="1168" | Find all messages in chats with exactly 2443 messages for which we have the permission to use. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| empty_msg | node & meta::empty_msg="3" | Find all messages in chats with exactly zero empty messages. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| empty_msg | node & meta::encrypted_msg="3" | Find all messages in chats with exactly zero encrypted messages. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| media_msg | node & meta::media_msg="3" | Chats containing a specific number of messages which originally had media (e.g. videos or pictures) attached. . This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| system_msg | node & meta::system_msg="3" | Find all chats that contain a specific number of system messages (such as "left the group"). This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |

# Message annotations

| **name** | example | |
|---|---|---|
| lang_source | lang_source="automatic" | Many messages have an annotation for "most_likely_lang". Some of those likelihoods were processed automatically, i.e. by means of statistical methods, others were annotated manually (mostly Romansh messages). The process is reflected in this field, options are "automatic" and "manual". |
| msg | msg="mediaQremoved" | Find messages which originally contained media such as pictures or videos that were removed. |
| msg_characters | msg_characters="1" | Find messages with a certain number of characters. . This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| msg_emojis | msg_emojis="3" | Find messages with a certain number of emojis. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| msg_id | msg_id="1273570" | Find messages with a specific ID. |
| msg_is_empty | msg_is_empty="true" | Find empty messages |
| msg_tokens | msg_tokens="1" | Find messages with a specific number of tokens. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| msg_type | msg_type="content" | Find messages that are not media messages or empty messages or messages without permission or technical messages (like "left the group"). Basically that means: normal messages written by humans. |
| msg_url | msg_url="#c=WUS&_q=bXNnX2lkPSIyOTQ0MjUi" | This is a technical field that is used to show one specific field. You cannot query it directly. Instead, the respective query is created when you click on the message ID in the chat display. |
| msg_vis | msg_vis="" | This field is mostly used for emojis, if you want to query them as emojis (as opposed to transcribed emojis like emojiQfaceThrowingAKiss). |

| | | |
|---|---|---|
| spk | spk="spk2963" | Find messages written by a specific informant. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| timestamp | timestamp="14 Jan à 13:52" | Find messages with a specific time stamp. Please keep in mind that the timestamp depends on the language used by the informant. This is an alphanumeric field, i.e. you cannot search for larger than or smaller than. |
| most_likely_lang | most_likely_lang="gsw" | Find messages which were annotated either by humans or by a computational linguistics tool as being most likely in a specific language. |

# Token annotations

| **name** | example | |
|---|---|---|
| gloss | gloss="viel" | Where messages have been normalized (i.e. "translated" into a standard variant and/or spelling), you can find this glossing or normalization here. |
| mftb_lem | mftb_lem="cln" | French messages that received Part of Speech treatment can be queried for the lemma assigned by the PoS tagger MElt. |
| mftb_pos | mftb_pos="NC" | French messages that received Part of Speech treatment can be queried for the PoS assigned by the tagger MElt. |
| pos | pos="PUN" | A generic Part of Speech annotation used for all languages points out features in common such as punctuation and emoticons |
| tt_lem | tt_lem="_UNKNOWN_" | German and Italian messages that received Part of Speech treatment can be queried for the lemma assigned by the PoS tagger TreeTagger. |
| tt_pos | tt_pos="NOM" | German and Italian messages that received Part of Speech treatment can be queried for the PoS assigned by the tagger Treetagger. |

# Demographic annotation

- Demographic information is attached to every message written by a specific informant.
- Some postal codes, cantons and cities are marked with an asterik in cases where we looked them up in lists. For example, if a communication partner left the field for the city and the canton empty but gave his postal code as 4144, we added the city as **Arlesheim* and the canton as **BL*.
- Also keep in mind that answers can be multiple, i.e. somebody can give "gsw, fra, ita" as their mothertongue if they are trilingual.
- To see the corresponding questions in all languages, please check the questionnaire.

| **name** | example | available values |
|---|---|---|

| | | |
|---|---|---|
| age_range | age_range="18-24" | 0-17, 18-24, 25-34, 35-49, 50-64, over 64, unknown |
| education | education="secondary school qualification" | university or polytechnic diploma, still in education, secondary school qualification, no indication, higher vocational education |
| flatrate | flatrate="yes" | yes, no |
| features | features="abbreviations,non standard,smileys,dialect,multiple languages" | non standard, multiple languages, smileys, dialect |
| gender | gender="unknown" | m, f |
| home_country | home_country="CH" | AT (Austria), CA (Canada), CH (Switzerland, CZ (Czech Republic), DE (Germany), FI (Finnland), FR (France), HN (Honduras), IT (Italy), LU (Luxembourg), PL (Poland) |
| home_postcode | home_postcode="1004" | |
| homelanguage | homelanguage="gsw" | deu (non-dialectal German, fra (French), eng (English), ita (Italian), roh (Romansh), und (undefined), frp (Francoprovençal), lmo (Lombard), gsw (dialectal German) |
| input_method | input_method="without correction" | with correction, with prediction, without correction, without prediction |
| message_rate | message_rate="21-50" | 0-5, 21-50, 51-100, 6-20, over 100 |
| mothertongue | mothertongue="gsw" | deu, eng, fra, ita, roh, und, frp, gsw, lmo (see above) |
| outsidelanguage | outsidelanguage="gsw,deu,fra,eng" | deu, eng, fra, ita, roh, und, frp, gsw, lmo (see above) |
| person_url | person_url="#c=WUSdemographics&_q=ZGVtb2dyYXBoaWNzX2lkPSIzNjQi&cl=0&cr=0&s=0&l=1" | n/a |
| school_canton | school_canton="*ZH" | |
| school_country | school_country="CH" | |
| school_postcode | school_postcode="4144" | |
| school_town | school_town="*Zürich" | |
| unemployed | unemployed="no" | yes, no |

| | | |
|---|---|---|
| work | work="still in education or training" | clerical or similar / craft, skilled trade and similar / elementary occupation / executive and managerial / facilities and machine operation, assembly / fisheries, agriculture, and forestry / home maker / no indication / other / service sector and sales / still in education or training / teaching or academic / technical or similar |
| work_country | work_country="CH" | AT, CA, CH, DE, FR, HN, IT (see above) |
| work_postcode | work_postcode="2000" | |

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

Permanent link:
**https://whatsup.linguistik.uzh.ch/02_browsing/04_queries/05_fields**

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