

# 2.2 Layers of information

WhatsApp messages are built up in a hierarchy: a chat contains messages that contain tokens that contain characters. A corpus of WhatsApp chats should allow for all these layers to be queried. Additionally, there is meta-data about the chats (e.g. number of messages) and about the messages (e.g. the timestamp when it was written) and about the informant (e.g. his/her age) and about the tokens (e.g. part of speech). This makes our corpus a rather challenging and complex endeavor.

These layers can nicely seen when browsing results from a query:

The screenshot shows a web interface for browsing WhatsApp chat data. At the top, the path is 'WUS\_ITA\_TT > chat138 (msg 20 - 22)'. Below this, there are several layers of data:

- spk (speaker):** A row with 'spk365', 'spk366', and 'spk365'.
- tok (tokens):** A row of tokens: 'Anke adesso se vuoi Aeh ho solo 10 percento di batteria xo Ah ecco'.
- token attributes:** A table with columns for each token and its grammatical category.
- message attributes:** A table with columns for 'tok', 'msg', 'msg\_id', 'msg\_type', 'most\_likely\_lang', 'msg\_tokens', 'spk', 'demographics\_id', 'gender', 'age\_range', 'mothertongue', 'home\_postcode', 'school\_postcode', and 'timestamp'.

At the bottom, there are navigation options: 'chat (context)', 'chat (complete)', and 'chat (message attributes)'.

In this example, you find the chat

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

Permanent link: [https://whatsup.linguistik.uzh.ch/02\\_browsing/02\\_layers?rev=1573036139](https://whatsup.linguistik.uzh.ch/02_browsing/02_layers?rev=1573036139)

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

