

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:

7 Path: WUS_ITA_TT > chat138 (msg 20 - 22) left context: 1 right context: 1

spk	spk365											spk366											spk365
tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	per cento	di	batteria	xo	Ah	ecco									

token attributes

tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	per cento	di	batteria	xo	Ah	ecco
gloss	anche	adesso	se	vuoi	Aeh	ho	solo	10	per cento	di	batteria	però	ah	ecco
tt_pos	ADV	ADV	PRO:refl	VER:pres	NOM	VER:pres	ADV	NUM	NOM	PRE	NOM	ADV	INT	ADV
tt_lem	anche	adesso	se	volere	_UNKNOWN_	avere	solo	@card@	per cento	di	batteria	però	ah	ecco

message attributes

tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	per cento	di	batteria	xo	Ah	ecco
msg	Anke adesso se vuoi				Aeh ho solo 10 per cento di batteria xo				Ah ecco					
msg_id	165379				165380				165381					
msg_type	content				content				content					
most_likely_lang	ita				ita				ita					
msg_tokens	4				8				2					
spk	spk365				spk366				spk365					
demographics_id	45				49				45					
gender	f				m				f					
age_range	18-24				25-34				18-24					
mothertongue	ita,imo				ita				ita,imo					
home_postcode	1004				3014				1004					
school_postcode					6500									
timestamp	30 mar 13:31				30 mar 13:32				30 mar 13:32					

chat (context)
chat (complete)

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

Permanent link: https://whatsup.linguistik.uzh.ch/02_browsing/02_layers?rev=1573035569

Last update: 2022/06/27 07:21

