

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 be seen when browsing results from a query:

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 percento di batteria xo Ah ecco																																																																																																																																																																																																																																														
□ token attributes																																																																																																																																																																																																																																														
<table border="1"> <tr> <td>tok</td><td>Anke</td><td>adesso</td><td>se</td><td>vuoi</td><td>Aeh</td><td>ho</td><td>solo</td><td>10</td><td>percento</td><td>di</td><td>batteria</td><td>xo</td><td>Ah</td><td>ecco</td></tr> <tr> <td>gloss</td><td>anche</td><td>adesso</td><td>se</td><td>vuoi</td><td>Aeh</td><td>ho</td><td>solo</td><td>10</td><td>percento</td><td>di</td><td>batteria</td><td>però</td><td>ah</td><td>ecco</td></tr> <tr> <td>tt_pos</td><td>ADV</td><td>ADV</td><td>PRO:refl</td><td>VER:pres</td><td>NOM</td><td>VER:pres</td><td>ADV</td><td>NUM</td><td>NOM</td><td>PRE</td><td>NOM</td><td>ADV</td><td>INT</td><td>ADV</td></tr> <tr> <td>tt_lem</td><td>anche</td><td>adesso</td><td>se</td><td>volere</td><td>_UNKNOWN_</td><td>avere</td><td>solo</td><td>@card@</td><td>percento</td><td>di</td><td>batteria</td><td>però</td><td>ah</td><td>ecco</td></tr> </table>																tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	percento	di	batteria	xo	Ah	ecco	gloss	anche	adesso	se	vuoi	Aeh	ho	solo	10	percento	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@	percento	di	batteria	però	ah	ecco																																																																																																																																																																			
tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	percento	di	batteria	xo	Ah	ecco																																																																																																																																																																																																																																
gloss	anche	adesso	se	vuoi	Aeh	ho	solo	10	percento	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@	percento	di	batteria	però	ah	ecco																																																																																																																																																																																																																																
□ message attributes																																																																																																																																																																																																																																														
<table border="1"> <tr> <td>tok</td><td>Anke</td><td>adesso</td><td>se</td><td>vuoi</td><td>Aeh</td><td>ho</td><td>solo</td><td>10</td><td>percento</td><td>di</td><td>batteria</td><td>xo</td><td>Ah</td><td>ecco</td></tr> <tr> <td>msg</td><td colspan="4">Anke adesso se vuoi</td><td colspan="6">Aeh ho solo 10 percento di batteria xo</td><td colspan="5">Ah ecco</td></tr> <tr> <td>msg_id</td><td colspan="4">165379</td><td colspan="6">165380</td><td colspan="5">165381</td></tr> <tr> <td>msg_type</td><td colspan="4">content</td><td colspan="6">content</td><td colspan="5">content</td></tr> <tr> <td>most_likely_lang</td><td colspan="4">ita</td><td colspan="6">ita</td><td colspan="5">ita</td></tr> <tr> <td>msg_tokens</td><td colspan="4">4</td><td colspan="6">8</td><td colspan="5">2</td></tr> <tr> <td>spk</td><td colspan="4">spk365</td><td colspan="6">spk366</td><td colspan="5">spk365</td></tr> <tr> <td>demographics_id</td><td colspan="4">45</td><td colspan="6">49</td><td colspan="5">45</td></tr> <tr> <td>gender</td><td colspan="4">f</td><td colspan="6">m</td><td colspan="5">f</td></tr> <tr> <td>age_range</td><td colspan="4">18-24</td><td colspan="6">25-34</td><td colspan="5">18-24</td></tr> <tr> <td>mother tongue</td><td colspan="4">ita,imo</td><td colspan="6">ita</td><td colspan="5">ita,imo</td></tr> <tr> <td>home_postcode</td><td colspan="4">1004</td><td colspan="6">3014</td><td colspan="5">1004</td></tr> <tr> <td>school_postcode</td><td colspan="4"></td><td colspan="6">6500</td><td colspan="5"></td></tr> <tr> <td>timestamp</td><td colspan="4" rowspan="3">30 mar 13:31</td><td colspan="6" rowspan="3">30 mar 13:32</td><td colspan="5" rowspan="3">30 mar 13:32</td></tr> </table>																tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	percento	di	batteria	xo	Ah	ecco	msg	Anke adesso se vuoi				Aeh ho solo 10 percento 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					mother tongue	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				
tok	Anke	adesso	se	vuoi	Aeh	ho	solo	10	percento	di	batteria	xo	Ah	ecco																																																																																																																																																																																																																																
msg	Anke adesso se vuoi				Aeh ho solo 10 percento 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																																																																																																																																																																																																																																			
mother tongue	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=1573035606



Last update: 2022/06/27 07:21