

What's up, Switzerland? Language, Individuals and Ideologies in Mobile Messaging

E. Stark¹, (T. Samardžić²)

¹URPP Language and Space, UZH, ²URPP Language and Space Lab, UZH.

The Project

- Funded by the Swiss National Science Foundation
- Project sum: CHF 1'787'280
- Project leader: Elisabeth Stark (estark@rom.uzh.ch)
- Four sub-projects; two post-docs, four docs
- Involved Universities: Zurich, Bern, Neuchâtel, Leipzig
- Project duration: 36 months (1/1/2016 – 31/12/2018) plus 14 months' extension

Two Overall Research Questions

- What do Swiss WhatsApp messages look like? What has changed overall between Swiss SMS and Swiss WhatsApp messages, and why (as regards linguistic structures, use of images in a broad sense, spelling, register-specific style, individualization vs. accommodation)?
- What is said / done by the individual users and the media in/on WhatsApp messages and chats, in relation to the findings for question 1?

Where we stand

- All doctoral students and the postdocs have finished their empirical analyses and are now writing up.
- (Post-)docs wrote 10 papers based on the corpus.
- CL-Team wrote three papers about the corpus and its data processing.
- 115 presentations were given by the team.
- Altogether, 39 publications with a link to the project / the corpus were published. Many more are in preparation.
- The corpus was used by around 30 people outside the project for their research.
- 136 articles about the project appeared in the printed press, radio or television until now.
- The project was also presented to a general public, e.g. at the Scientifica (UZH, 2018: A. Göhring; 2019: E. Stark), at the Kinderuniversität 2017 (UZH, Ch. Siever) or at the Wissenschaftsfestival 2018 (UZH, Ch. Dürscheid).

Some selected publications

- Dürscheid, Christa/Siever, Christina M. (2017): Jenseits des Alphabets. Kommunikation mit Emojis. Zeitschrift für Germanistische Linguistik 45/2, 256–285.
- Felder, Samuel (in prep.): Individuelle sprachliche Variation in WhatsApp-Chats. Stil, Akkommodation und *real-time change* (Dissertation).
- Luseti, M., Ruzsics, T., Göhring, A., Samardžić, T., and Stark, E. (2018): Encoder-Decoder Methods for Text Normalization. In Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018), 18–28. Santa Fe, New Mexico, USA.
- Natale Silvia (2019): Information selection e information packaging. Alcuni spunti sull'organizzazione dell'informazione in testi narrativi orali in italiano e francese. *Studia linguistica romanica* (1).
- Petitjean, Cécile/Morel, Etienne (2017): "Hahaha": Laughter as a Resource to Manage WhatsApp Conversations. *Journal of Pragmatics* 110, 1-19.
- Ruzsics, Tatiana; Luseti, Massimo; Göhring, Anne; Samardžić, Tanja; Stark, Elisabeth (accepted): Neural Text Normalization with Adapted Decoding and PoS Features. *Natural Language Engineering*.
- Siebenhaar, Beat (2018): Funktionen von Emojis und Altersabhängigkeit ihres Gebrauchs in der WhatsApp-Kommunikation. In: a. Ziegler (ed.): *Jugendsprachen. Aktuelle Perspektiven internationaler Forschung*. Berlin: De Gruyter.
- Thurlow, Crispin (in press): Mediatizing sex: Sexting and/as digital discourse. In K. Hall & R. Barrett (eds). *The Oxford Handbook of Language and Sexuality*. New York: Oxford University Press.
- Ueberwasser, Simone/Stark, Elisabeth (2017): What's up, Switzerland? A corpus-based research project in a multilingual country. *Linguistik online* 84/5, 105-126.

The Corpus

- Planned publication (open access): end of 2019
- 617 Chats
- 763,650 messages with text
- 5,543,692 tokens
- Texts from 945 participants, 426 with demographics
- Multilingual (German, French, Italian, Romansh)

Language	Chats	Messages	Tokens
Non-dialectal German	93	81,456	625,419
Swiss German	275	506,984	3,611,033
French	141	197,255	1,397,375
Italian	87	42,559	293,567
Romansh	77	29,094	283,909

Data processing

- Masking of non-consented messages
- Anonymization
- Mapping of emoji codes to Unicode and *emojiQdescription*
- Tokenization
- Removal of duplicate chats
- Language identification for chats and messages
- PoS annotations and normalization for the French and non-dialectal German corpus
- PoS annotations for the Italian corpus
- Manual normalization of parts of the Swiss German corpus
- PoS annotation and normalization for the whole dialectal German corpus (using Neural encoder-decoder models)

Some findings

- Accommodation can be an important factor influencing changes on different linguistic levels and long-term developments of individuals' language use may be influenced/triggered by cases of short-term accommodation (Felder (in prep)).

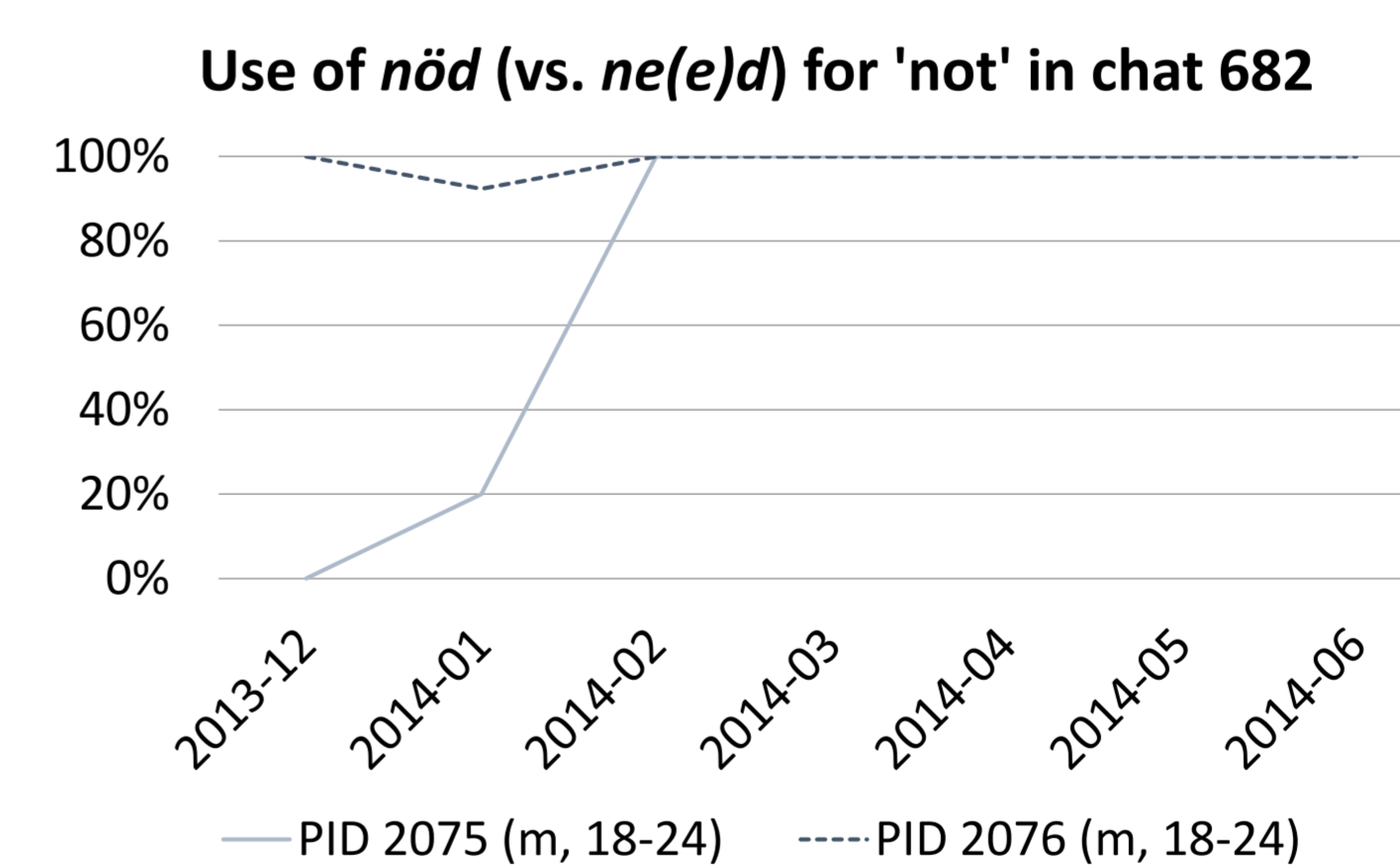


Fig. 1: Informant 2075 shows a rapid convergence towards the same language use as informant 2076.

- In our corpus of *WhatsApp* messages as well as in our additional corpus of *Instagram* comments, different emojis are used most frequently (slightly modified from Dürscheid/Siever (2017)).

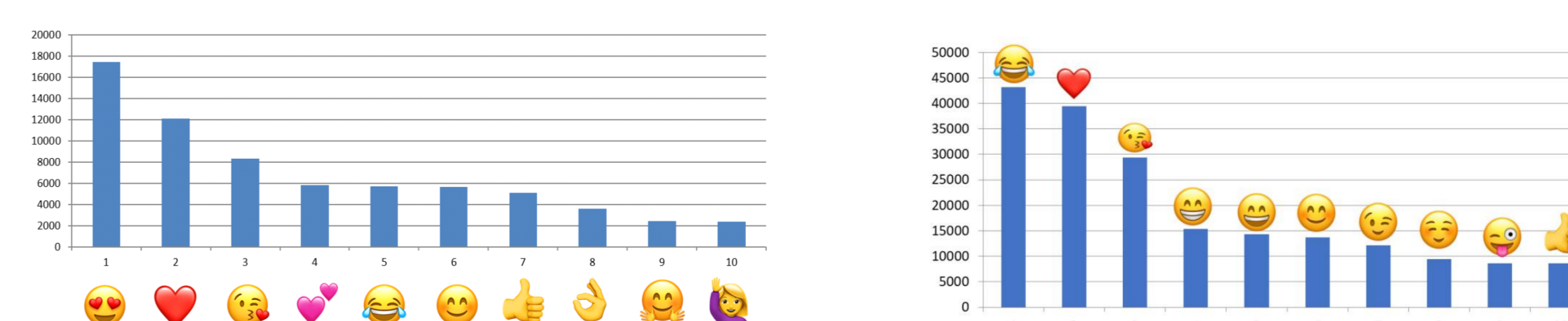


Fig.2: Top ten emojis in the *Instagram* (left) and *WhatsApp* corpus (right).