RASK 52
AUTUMN 2020

Annotating Emoticons and Emojis in a German-Danish Social Media Corpus for Hate Speech Research

ECKHARD BICK

Department of Language and Communication, University of Southern Denmark

ABSTRACT
This paper presents and evaluates the emoticon/emoji annotation of a large German-Danish Social Media corpus for hate speech research. Overall tagging was more accurate for emojis (98-99%) than for text emoticons (91-92%), and slightly better for German than for Danish. We discuss problems and strategies involved in the recognition and linguistic annotation of emoticons and emojis, and show how an emoticon classification system can be used to highlight interesting differences between German and Danish Twitter, as well as between the background corpus on the one hand and tweets targeting the immigrant/refugee minorities on the other. Using concrete examples, we illustrate how the annotation facilitates corpus inspection and how certain emoticon types (e.g. 'wink' and 'skeptical') can help to identify otherwise inaccessible examples of non-direct hate speech. Finally, we use emoji-informed word embedding to investigate the emotional content of equivalent immigration key words in German and Danish.

1 Introduction

In the face of growing concerns about online manifestations of Hate Speech (HS) against minorities (e.g. Foxman & Wolf 2013), social media are trying to identify and block HS content with automatic means. The resulting conflict between freedom of speech on the one hand and minority protection on the other (Herz & Molnar 2012) has created a need for reliable data, actionable definitions and linguistic research in the area. While trigger words for outright HS (such as slurs) are relatively easy to identify, it is much harder to recognize covert use of HS manifesting with patterns such as generalization, othering, irony or sarcasm. In this context, emoticons and emojis, common in computer-mediated communication (CMC), may help to flag negative sentiment or even specific emotional reactions to events or people (Hauthal et al. 2019) as well as non-literal meaning of utterances. This paper presents an annotation scheme for such CMC symbols, evaluates their prevalence and usage in a German-Danish Twitter corpus and discusses how such annotation can facilitate corpus inspection and help to identify research-relevant content, as well as quantify keyword-associated sentiments.

The corpus was compiled for a three-year HS research project (XPEROHS, Baumgarten et al. 2019) investigating the (linguistic) expression and perception of online hate speech with a

eckhard.bick@mail.dk

particular focus on immigrant and refugee minorities in Denmark and Germany. Apart from linguistic studies, the data has also been used to provide graded HS examples for the project's empirical work on HS perception (questionnaires and HS stimuli for eliciting physiological reactions). The corpus is a monitor corpus (starting December 2017) with continuously collected data. In order to avoid excluding unthought-of HS variants, broadest possible coverage was aimed for, by using high-frequency function words as search terms for the Twitter API (e.g. *und/og* [and], *oder/eller* [or], *der-die-das/den-det* [the], *er-sie-es/han-hun* [he-she-it], *ist/er* [is]). The unabridged and bilingual character of the corpus sets it apart from monolingual, hashtag-driven or author-based data sets such as Kratzke's (2017) work on German parliamentary elections.

All in all, the corpus contains about 4 billion words, 3,700 million in the German section and 270 million for Danish. Both sections were annotated linguistically at both the morphosyntactic and semantic levels, using Constraint Grammar methodology (Bick & Didriksen 2015; Bick 2020). For the emoticon/emoji annotation discussed in this paper, we added specific modules (program sections) to the preprocessor and morphological analyzer stages of these parsers (GerGram[1] and DanGram[2] for German and Danish, respectively), providing a tokenization and classification scheme that would allow the new tokens to fit into syntactic sentence trees without negatively interfering with existing parsing rules.

2 Emoticons versus emojis

Emoticons and emojis (sometimes called graphic emoticons) are not always clearly distinguished, and the former are often automatically transformed into (a subset of) the latter by texting tools such as smartphones or email editors. However, there are both conceptual and formal differences, the latter being crucial for tokenization. Emoticons are pictorial representations of facial expressions, using punctuation symbols and, to a lesser degree, a few numbers and characters, for instance :-) -- the classical lying-down smiling face with eyes ':', nose '-' and mouth ')'. Emojis are a newer phenomenon, "hieroglyphical" in nature, with a Japanese origin and inspired by a writing system using ideograms rather than a (phonetic) alphabet. In fact, the term 'emoji' roughly translates as 'icon', with no relation to the English word 'emotion', and the universe of emojis contains not only the (emotion-bearing) smiley faces, but also a wide range of objects, actions, places etc. For want of an established neutral term ('emoglyph' could be a suggestion), we will here use 'emoticon' as an umbrella term for both emoticons and emojis, using - where necessary - 'text emoticon' to distinguish the hyponym from the hypernym.

While some emojis can replace words in a text (in our corpus flag emojis were used instead of country names), it can be argued that most emoji usage constitutes an additional, non-textual modality, with its own syntax:

1 https://visl.sdu.dk/de/parsing/automatic/
2 https://visl.sdu.dk/da/parsing/automatic/

> Emoji can be used as a supplemental modality to clarify the intended sense of an ambiguous message, attach sentiment to a message, or subvert the original meaning of the text entirely in ways a word could not. Emoji carry meaning on their own, and possess compositionality allowing for more nuanced semantics through multi-emoji phrases. (Pereira-Kohatsu 2019).

Still, in most cases emoticons correlate with the content or sentiment of the preceding or embedding text, allowing machine-learned models to predict one from the other with a certain confidence (Cappallo et al. 2019). Therefore, corpus-linguistically, emoticons fulfill a double function: On the one hand, qualitatively, they can help assess ambiguity, sarcasm and skepticism. On the other hand, they can help searching for pejorative language or quantify sentiments associated with search terms found in the same tweet.

2.1 Tokenization: Recognizing and delimiting emoticons and emojis

In terms of (automatic) corpus annotation, emoticons are an interesting topic for two very different reasons: First, they are an important element of a set of ”colloquial” traits setting CMC discourse apart from ordinary written text (Beißwenger et al. 2016), and as such present a formidable obstacle for a parser that has not been designed to handle them. All other things being equal, text emoticons will end up split into punctuation ”atoms”, and emoji strings as out-of-vocabulary (OOV) foreign nouns. The former can lead to sentence discontinuities and structural parsing errors, while the extra constituents spawned by the latter may affect NP cohesion, interfere with uniqueness rules, or even mask adjacent words into OOVs in the absence of a separating space character. The second reason for taking emoticons seriously in HS-oriented corpus linguistics is the obvious one - their emotional content. Thus, if correctly recognized and annotated, emoticons can help to decide the degree of HS of a given utterance, or even help to search for new hateful content.

Several parser adaptations were necessary to handle this new class, the first step being (a) pattern-based recognition of text emoticons as character/punctuation strings and (b) a separation and marking of emojis based on Unicode blocks.

(a1) simple smileys: ([:\-,;o]+)
(a2) eye and letter smileys: (>?[:;>\}%]\'?-?
*[coXx8]?[\)\|]?[\)\(\]\[\}\{\/\\\|<&\#\$\@3bcDJdLPoOopþSX]|X\()
(a3) heart smileys: ([<:](- *)?3+)
(b) \x{1F300}-\x{1FA9F}
\x{1F1E6}-\x{1F1FF}

In both cases context can be relevant. Thus, text emoticon ”chaining” has to stop before true punctuation and stay clear of real (e.g. ordinal) numbers (3., 8.). For emojis, a distinction has to be made between multi-character emojis on the one hand (e.g. the composite emoji meaning ’family’) and the chaining of individual emojis on the other (e.g. skeptical face + angry face).

2. 2 Annotation and classification

Once tokenized, the individual emoticons are then classified and annotated. First, in order to be assimilated into the general parser annotation and to become part of the syntactic tree structure for a given utterance, morphosyntactic mark-up is needed. For smiley emoticons, syntactic function can mostly be said to be adverbial, "commenting" on a preceding or following chunk of text by assigning sentiment in much the same way adverbs would (e.g. 'lamentably', 'luckily', 'sadly' etc.). Choosing ADV (adverb) as part-of-speech (POS) for emoticons will not only trigger adverbial function tagging by the parser, but is also the POS least likely to interfere with existing structural rules in its grammar, because adverbs in both German and Danish have fewer positional constraints than other POS.

Though emojis can obviously denote concepts that are nominal or even verbal in nature, the vast majority, at least in our data (cp. section 3), consists of faces and gesture symbols with emotional content, and is used much like text emoticons (a, b). In fact, for the most frequent emojis, both social networks and smartphone operating systems have published one-on-one correspondences with emoticons. Therefore, emojis are morphosyntactically annotated as adverbs/adverbials, too (a, b below). A POS exception are flag emojis (c), that get tagged as proper nouns, because they are sometimes used instead of country names in the HS corpus.

(a) *Diese ganze Anti-Flüchtlings-Polemik geht mir auf die Nerven.* 😠 (All this refugee bashing drives me insane)
(b) *Die nächste Moslemlawine, die über unser Land hinwegrollt. .... Es ist nichts geringeres als ein Völkermord auf Raten.* 😠 (The next Muslim avalanche hitting our country ... Nothing less than a genocide by installment)
(c) *er ist gut für ein starkes* 🇫🇷 *während* 🇩🇪 *vor die Hunde geht* (he is good for a strong France, while Germany goes down the drain)

At the syntactic level, among the emojis frequent in our corpus, skin-colour emojis (a class of 5 skin tones) also constitute an exception, because they are used to post-modify other emojis - in particular, but not exclusively, face and person emojis. In these cases, the adject function and close dependency attachment are used rather than clause-level adverbial tagging (cp. section 5).

2.3 Semantic/sentiment classification

For semantic classification, following Hauthal et al. (2019), we group emoticons into emotional (sentiment) umbrella classes, e.g. "emo-happy", "emo-love", "emo-sad", "emo-angry" etc. However, Hauthal's 6+1 scheme was modified by splitting his 'joy' class into 'happy' and 'laugh', and by adding 'wink' and 'skeptical' as individual classes[3]. Because these classes lump together emoticon variants having similar meanings, they are assigned the feature slot of "lemma" (which is otherwise used to bundle inflectional paradigms under one dictionary entry).

3 With their implication of relativized truth value, the latter two classes were deemed to be particularly important for research into non-literal expressions of hate speech.

This decision aligns with the use of the lemma slot for canonical forms bundling non-standard spellings and mistyped word forms – a use also found abundantly in CMC corpora.

Unlike text emoticons, emojis are given individual lemmas (e.g. ”emo-gesture-Left-Facing-Fist”), following a technique of ”emoji translation” suggested and used by several participants in the HS shared task in SemEval-2019 (Basile et al. 2019). For emojis with emotional content, the first part of the lemma will often indicate one of the 9 emotional umbrella classes (e.g. ”emo-laugh-Face-With-Tears-of-Joy”). In theory, individual lemmas are superfluous, because the emoji itself (the ”word form”) contains the same information. However, on a computer keyboard, emojis are more difficult to type than text emoticons, whereas the explicit textual lemma with a class prefix facilitates corpus searches.

Table 1 lists the different emoticon classes and all the emoticons found in the corpus with a frequency > 1%, in some cases followed by an (incomplete) list of rarer examples (other).

Table 1: Emoticon classes (examples > 1%)

EMOTICON CLASS	TYPES BY RANK (OTHER: < 1%)
emo-happy	😎 :) 😊😈😇 :-) 🙃🙂😌:3 :)) other: :-)) :]:-]:}:-]
emo-love	😍😘🤗<3 ☺ 😋💋🤤🤩♡😙😙
emo-laugh	😂🤣😁😅😆😀😄😃
emo-sad	😭😢 :(😥😞 :-(😴😰😪😓🙀 other: :< :-[:@ :[:-@
emo-horror	😱😨😩😫😟😦
emo-angry	😡😠💩😤👿💀😣 other: ☠ X(
emo-surprise	😳🤭😲😖😮😯
emo-wink	😉 ;) 😜😏;-) 🤪😒😝🤑
emo-skeptical	🤔🙄🧐🤨🤬😬😔🤓😕😐😟😑🤥
emo-other (emoticons only)	;-)) ;)) <333 (intensifiers) ;-(;((combination wink+sad) 8D

As can be seen, text emoticons only make it above the 1%-threshold for four classes, occupying rank 2 for ’happy’ and ’wink’, rank 3 for ’sad’ and rank 4 for ’love’). The 10th class, ’emo-other’, functions as a catch-all for rarer emoticons that matched the general pattern constraints and thus were identified as emoticons but had not been classified into one of the nine ‘sentiment’ classes yet. In connection with iterated annotation runs, emoticons can be moved from this ”waste-bin” class to a real class. The double-bracket wink emoticons (’;))’ and ’;-))’), for example, were not covered at evaluation time, but could be read as an emphasized version of the single-bracket standard ’;)’.

3 Statistical evaluation

In order to get an idea of the quantitative importance of different emoticons/emojis and emo-classes, as well as their possible functional and language differences, we performed a statistical evaluation of the annotated corpus.

3.1 Prevalence of emoglyphs

In the unfiltered Twitter corpus, 13-17% of tweets contain either text emoticons or emojis or both, a bit more for German than for Danish (table 2) and much more than the 4% reported as an average for 13 European languages by Novak et al. (2015)[4]. Of those tweets containing emoticons/emojis, many have more than one, the average density being slightly higher for Danish (1.7) than for German (1.2). The per-word density of emoticons was also higher for Danish (21.4 / K words) than for German (14.8 / K words), a difference also noted by Ljubešic & Fišer (2016) in their cross-language study on emoji usage.

Table 2: Emoticon prevalence

EMOTICON/EMOJI DENSITY	TW-DE		TW-DA	
	all	immi	all	immi
words	1,700 M	36 M	140 M	2.6 M
tweets	119.5 M	1.7 M	13.1 M	106 K
w/emoticons	2.3%	0.4%	2.0%	0.6%
w/emojis	14.9%	4.5%	11.4%	5.9%
w/emo (all)	17.2%	4.9%	13.4%	6.5%
emoglyphs	25.2 M	294 K	3.0 M	12 K
emoticons	10.7%	7.3%	8.7%	6.6%
emojis	89.3%	92.7%	91.4%	93.5%
emo/emotweet	1.2	3.5	1.7	1.74
emo/K words	14.8	8.2	21.4	4.6

The balance between emojis and text emoticons was roughly 10:1 in favour of the former, possibly because of automatic conversion of common text emoticons by texting systems. Interestingly, sub-corpora keyword-filtered for the immigrant/minority topic[5] had a considerably lower prevalence of emoticons/emojis (per 1000 words), albeit with the same balance between emoticons and emojis. The reason for this is likely a different distribution of types, with an un-

4 The difference is likely not language-dependent, but rather caused by a general development towards increased use of emojis over time.

5 The keyword lists were roughly equivalent and of similar size for the two languages. However, unlike the more constrained noun-based immigration filter used in sections 3.2. and 3.3., this filter was designed for maximum recall in corpus inspection (rather than specificity for the topic), and included typical verbs and general pejorative terms, yielding a much larger filtering space (800+ words).

der-representation of the ordinary happy and love smileys (cp. sections 3.2 and 3.3), in favour of otherwise rarer, more aggressive smileys ('angry', 'skeptical'). This is in line with Novak et al.'s (2015) finding that tweets with emojis have a more positive sentiment mean than those without, suggesting that the negative sentiment found in the refugee discourse would lead to fewer (but angrier) emojis. In addition, emoticons in that subdomain are concentrated in a smaller percentage of tweets. Even given that these tweets at the same time are longer than those in the background corpus, there are still relatively many emoticons in those tweets that do have those emoticons (3.5 per tweet for German).

3.2 Distribution of sentiment classes

One of the main reasons for annotating emoticons was that we expected them to be potential flags of strong opinions capable of expressing irony, sarcasm, a demeaning or hateful stance in connection with otherwise innocent remarks. Thus, Wang & Castanon (2015) found that removing emoticons cut the number of tweets marked as sentiment-polar by half. This assumption can be corroborated not only by inspection, but also through a frequency break-down of the 9 sentiment tweet classes classes used in our annotation scheme (Table 4). For comparison purposes, we created comparable immigrant subcorpora by prefiltering for tweets containing a list of immigration-related marker-morphemes[6] that were roughly equivalent in both languages.

Table 4: Relative distribution of the 9 sentiment tweet classes

EMOGLYPH CLASS	TW-DE		TW-DA	
	all	immigrant subcorpus	all	immigrant subcorpus
emo-happy	**15.9**	10.4	**23.8**	16.1
emo-love	**16.6**	4.6	**9.5**	6.6
emo-laugh	31.0	**39.2**	31.2	27.6
emo-sad	6.1	3.6	8.6	8.6
emo-horror	2.8	2.6	2.5	1.8
emo-angry	2.8	**8.3**	2.0	**5.9**
emo-surprise	2.0	2.1	2.4	3.6
emo-wink	11.7	11.9	12.0	11.5
emo-skeptical	11.0	17.2	7.9	**18.2**
	99.9	99.9	99.9	99.9

As can be seen, for both languages there are indeed striking differences between the neutral background corpus and the immigrant minority subsections. First of all, the dominating gen-

6 The lists, expressed as regular expressions, were (a) ".*(afrikan|arab|asyl|flücht|immigr|einwander|islam|migrant|m[uo]sl[ei]m|musel|neger|nigg|alman|kanac?k|nafri|syrier|türk|ausländer).*" and (b) ".*(afrikan|arab|asyl|flygt|immigr|indvandr|migrant|islam|m[uo]sl[ei]m|muhamedaner|neger|nigg|nydansker|perker|syrer|tyrk|udlænding).*", for German and Danish, respectively

eral class of 'emo-happy', that includes e.g. the classical :) smiley, is 1/3 lower in the immigrant discourse. The 'love' emoji is even more underrepresented, especially in the German data. Conversely, statements about fugitives, foreigners and immigrants exhibit a higher prevalence of the aggressive 'angry' emojis. In absolute terms, the largest overweight is found for 'skeptical' emoglyphs. Linguistically this class and the more evenly distributed 'wink'-class are of special interest, because they can help to locate non-literal meaning and truth value inversions.

3.3 Most popular emoticons/emojis

Table 3 presents individual emoticon frequencies for the German Twitter corpus and for an immigration-filtered subsection (cp. chapter 3.2.), with strong overweights in boldface.

Table 3: Top ranking emoticons/emojis in the German Twitter corpus (% of all emoticons)

EMOTICON/EMOJI	FREQUENCY (%) IMMIGRATION SUBCORPUS	FREQUENCY (%) ALL TWITTER
😂 laugh & tears	**13.5**	9.0
🤔 skeptical	**4.5**	2.2
🤣 ROFL	**3.9**	2.0
skin colour suffix emoji 🏻 light 🏼 medium light 🏽 medium 🏾 medium dark 🏿 dark	3.9 46.5 29.5 10.2 6.6 7.1	3.2 51.6 33.9 9.4 2.9 2.2
😉 winking	2.8	2.4
😡 angry pouting	**2.4**	0.6
♂ male suffix emoji	**1.9**	1.0
👍 thumbs up	2.3	2.3
🤮 vomiting	**1.9**	0.4
🙄 skeptical rolling eyes	**1.8**	1.2
👎 thumbs down	**1.6**	0.2
🤦‍♂️ face palming	**1.5**	0.6
😎 happy sunglasses	1.5	1.1
😁 laugh smiling eyes	1.4	1.5
:), :-) happy emoticon	1.4	**2.5**
☆ star	1.1	0.6
🤷‍♀️ person shrugging	1.0	0.8
♀ female suffix emoji	1.0	0.9

❤ red heart	1.0	**3.3**
;), ;-) wink emoticon	1.3	1.6
💥 collision (conflict)	**0.9**	0.2
😅 laugh with sweat	0.8	**1.4**
😊 smiling eyes	0.8	**1.5**
🙈 no-see monkey	0.8	1.1
😍 smiling heart eyes	0.8	**2.3**
😠 angry face	**0.8**	0.2
👏 clapping	0.7	0.5
😱 horror screaming	0.7	0.6
👉 pointing finger	0.7	0.5
🤢 nauseated face	**0.6**	0.2
😜 winking face	0.6	0.5[7]
…		
😘 blowing a kiss	0.4	1.7
😭 sad crying	0.5	1.2
🍀 four-leaf clover	0.2	0.9
🤗 love hugging		0.9
<3 (heart emoticon)	0.2	0.7

Reminiscent of a Zipf curve, there a few dominant symbols, followed by a shallow tail of more evenly spread items. The most important sentiments for the minority subcorpus were those of ridicule ('laugh & tears') and skepticism ('skeptical face'). Together with the related 'ROFL' ('rolling on the floor laughing'), the former stand at 17.4%, while the latter score 7.6% together with 'rolling eyes' and the standard 'wink' symbols. Interestingly, many positive emotions (happiness, smiling, kissing and hearts) were much rarer in this domain than in the background corpus, suggesting that the immigrant discourse does not invite statements about your being happy, friendly or in love, but favours opinion expressions (ridicule, anger, thumbs-down) and truth assessments (skepticism). The overweight of the 'vomiting' and 'nausea' emojis in the immigrant domain can be explained by the negative German idiom *'zum Kotzen'* ("nauseating").

Somewhat surprisingly, classical happy smileys, textual or not, while more common in the background corpus, seem to be 'leaking' their meaning to newer, more specific symbols, e.g.

7 Although meaning-wise related, the "winking face" emoji is distinct from the "winking" emoji, both graphically and in terms of ISO code, and therefore gets its own statistics here.

'happy sunglasses' or 'smiling eyes'. Gender and skin colour emojis are used as suffixes for people emojis, and therefore ranked relatively high[8]. In a hate speech context, skin tone emojis are interesting, because they can be used as a vehicle for *othering*: light=we, dark=you (foreigners). 'Male' was overrepresented due to the many comments directed at male immigrants and their behaviour.

4 Non-direct hate speech

Direct HS markers such as slurs and abusive language are easy to search for, but they cover only part of the phenomenon, and many hateful tweets will not even make it into the corpus in the first place, because Twitter's HS filters can identify (and remove) them just as easily. Non-direct HS, on the other hand, is difficult to identify automatically or to inspect systematically in a corpus, because it lacks the necessary trigger tokens, while negative sentiment is here expressed solely by indirect means, such as irony, sarcasm and metaphor. In these cases, wink- and skeptical emoticons are among the few surface markers available for an otherwise very difficult task - not because they necessarily over-correlate with HS (like emo-wink does not!), but because they can turn seemingly neutral or positive statements into negative ones.

The GermEval shared task on the automatic identification of offensive language (Wiegand et al. 2018) also highlights this difference. Key-word-based, explicit offense was mostly detected, while keyword-lacking offense, being implicit, was mostly missed even by the best performing systems. However, missing keywords can be compensated for by using emoticons, either as direct sentiment carriers or by flagging non-literal meaning. Thus, Prereira-Kohatsu et al. (2019) demonstrated the semantic similarity, in terms of word embeddings, between emojis and emotion-carrying trigger words, whereas Felbo et al. (2019) achieved 69-75% accuracy in sarcasm detection, using emojis with an ML classifier[9].

The German[10] examples (a-e below) demonstrate this mechanism. The angry emoji in (a) simply aligns with the slur term 'Musel', and either would be enough to doubt (negate) the literal meaning of the proposal. However, though it nicely fits the "living-at-our-expense" narrative, finding and interpreting (b) as hateful in a billion-word corpus would be difficult with ordinary annotation alone. Even if "*auf großem Fuß*" gets recognized as a fixed expression, it risks getting flagged as positive in isolation. Together with an angry-face emoji and a left-facing fist, the meaning is clear and the example is likely to find its way into a concordance. Finally, the textual sadness emoticon in (c1), as well as the prototypical wink smileys in (c2) and (c3), are a means to underline (and identify in corpus searches) the intended ironic (c1-2) or sarcastic (c3) interpretation of the utterances.

8 Of course, it is an artifact of tokenization that the tone emojis are visibly separately - in an actual tweet they would be fused with their head emoji, and the author would not compose the symbol, but simply choose between e.g. a light person and a dark person.

9 A demo of the system, DeepMoji, is accessible at: https://deepmoji.mit.edu/

10 For space reasons, we will focus on German examples exclusively, deeming a more thorough discussion to be more important than an extra set of Danish examples.

(a) *Aber ja , holt noch mehr Musels nach Deutschland rein ...* 😠 (But yes, go get even more Muslims to Germany)
(b) *Rentner sammeln Flaschen und Flüchtlinge leben auf großem Fuß* 😡👉 (Pensioners collect bottles, refugees live rich)
(c1) *Unsere fiese, unmenschliche Regierung ist einfach zu gemein zu den armen Flüchtlingen. :(* (Our nasty, inhumane government is simply sooo mean to the poor refugees)
(c2) *Für die «Schutzsuchenden» gibt HH ca. 1 Milliarde pro Jahr aus .* 😏 *Man muss halt Prioritäten setzen .* 😏 (For the "refuge-seekers" Hamburg spends 1 billion a year. Well, you need to have priorities)
(c3) *Die Asylanten sind fachkompetent, ganz lieb und es sind keine Kriminelle bzw. Terroristen mit dabei. Wurde uns so versprochen* 😉 (The asylum-seekers are professionally educated, nice people. No criminals or terrorists among them. Or so we were promised.)

5 Emoticon syntax

Tagging emoticons/emojis as adverbial utterance modifiers is a convenient compromise, but it is rooted in the usage of traditional text emoticons; by contrast, current emoji usage can exhibit a much more complex syntax, the description and implementation of which in our parsing scheme is ongoing work. Thus, both (a) and (b) are post-modifiers, but only the former adverbially comments on the whole utterance - the latter is in fact a post-nominal attribute (of *Speichelleckerin* and *EU,* respectively). In (c), emojis are used as prefix morphemes, i.e. inside the intended words, '*Scheißreligionen*' (c1) and '*Ziegenficker*' or '*Kamelficker*' (c2)[11]. Another syntactic feature, repetition for emphasis, is rarely used at the word level in German and Danish, but occurs a lot with emojis, e.g. (c2) and (e3)

(a) *Ich krieg das kalte Kotzen!* 😠 [this nauseates me]
(b1) *diese peinliche Moslem Speichelleckerin* 😠 [this embarrasing Muslim-spitlicker woman]
(b2) *Die EU* 😠 *maßt sich an, darüber zu entscheiden, wie wir #Kindergeld für Migranten ...* [The EU presumes to decide how we [pay] child benefits]
(c1) 💩*Religionen* [shitty religions]
(c2) *schickt sie in die Wüste die*🐐🐪*ficker* 🎺 🍌 🍌 🥁 🎪 👌 👌 [send them into the desert, those goat/camel-fuckers]

More and more, emojis are also used as words themselves. Thus 'DE' in (d1), is rendered as a German flag and means 'Germany', while the pile-of-shit emoji fills the slot of '*Scheiße*' in the idiom '*in der Scheiße stecken*' (be in big trouble). The headscarf-woman emoji in (d1) means 'muslim/muslima' and is - like the prefixes in (c) - a method to avoid explicit slurs and thereby the attention of HS filtering robots. By linking emojis together like Chinese ideograms, even com-

11 Note that the co-ordination of prefixes in the latter is unique to emoji syntax and would not be possible in regular text.

plete syntactic trees (predications) can be built. Thus, the emoji string in (e1) can be ‘translated as “entrance to Germany forbidden”(👉🇩🇪⛔), (e2) means “beat (👊) groups of men (🚶 🚶) with dark skin (🏿)”, while (f), finally, is a case of mixed usage, with both ad-nominal attributes (🐪), adverbial modifiers (🤮😡🤢) and one as-is ideogram (🧕).

(d1) *wie tief* 🇩🇪 *in der* 💩 *steckt* (how fucked-up the German situation is)
(d2) *Nein,* 🧕 *die gehören ausnahmslos erschossen !* (No, those headscarfies should be executed)
(e1) *schmeißt die Illegalen raus* 👉🇩🇪⛔(throw out the illegal [immigrants])
(e2) *GEHT AUF DIE WILDEN MUSLIM. VERGEW.*[12] *LOS, DIE IM RUDEL JAGEN* ❗🚶🏿🚶🏿👊 (attack the barbarian muslim rapists, hunt them in packs)
(e3) *Geht alle heim In Syrien gibt es keinen Krieg mehr* 🔙🚶🏿🚶🏿🔙🔙🔙👊👊👊 (Go home, all of you, there is no war any more in Syria)
(f) *wir sollten sie aus unserem Land jagen..die Kanaken*🐪🧕🤮😡🤢 (we should throw them out of the country, those apes)

6 Cross-language sentiment comparison

Statistical links between certain emoticons and key terms in the immigration discourse can be used to quantitatively assess differences in sentiment between these terms and between related terms across languages. One way to establish these links is semantic similarity expressed as vector distances between so-called word embeddings. In this approach, the semantics of a word is approximated through the context words it is ‘embedded’ in, and word vectors are computed in a multi-dimensional space where dimensions are co-occurring words, with the values of 0 or 1. In the CBOW (continuous bag-of-words) model used here, co-occurrence means ‘occurring together’ anywhere in the same tweet. As our input data we used lemmatized text, stripped of non-inflecting word classes (i.e. using only nouns, verbs, adjectives and proper nouns); the model was trained using the word2vec neural-net method (Mikolov et al. 2013) and the TensorFlow suite (Abadi et al. 2016)[13]. Once trained, the model outputs vector similarities between two given lemmas, where 1 means a complete synonym (100%), and 0 means no similarity. Obviously, emoticon lemmas and noun lemmas cannot be complete synonyms, but they can share typical contexts, with embedding similarities of up to 35-45%. Significant correlations are typically above the 25% threshold, whereas below 15% similarity correlations are only useful in relative terms, if at all. Adjectives can provide a kind of calibration. The model assigned the following vector similarities for three negative German adjectives:

Vomiting emoji: *schlecht* [bad] 25.7, *böse* [evil] 27.8, *kriminell* [criminal] 43.3
Angry emoticon: *schlecht* [bad] 19.7, *böse* [evil] 24.8, *kriminell* [criminal] 26.5

12 Short for "muslimischen Vergewaltiger"
13 Following an extensive experimentation phase, the following parameters were chosen for both languages: dimensions = 100, iterations = 10, vocabulary = 100,000

Fig. 1 shows, for German/Danish noun pairs, the highest relatedness (in %) with a negative-sentiment emoticon class (the leftmost two subcolumns). Since this percentage can represent different emoticon classes for different nouns (cp. Table 5), similarities for the emo-*angry* class are provided for standardization (the rightmost two subcolumns).

The three immigration implying terms *immigrant, refugee* and *asylum-seeker* draw more emoticon-negativity for German than for Danish, while the religion-terms *Muslim* and *Islam* are more negative in Danish, implying a difference in focus beween the two countries. Among the slur terms, the Danish *perker* [foreigner_slur] appears to be more negative than the German equivalent *Kanacke,* while the German *Musel* [Muslim_slur] is worse than *muhamedaner* in Danish. The term *nazi* tops the list for both languages for all combinations, with the exception of the maxneg column for German *Musel*. Interestingly, and in contrast with ethnic and religious slurs, *nazi* seems to be morphing into a general, target-independent slur. Thus, in both

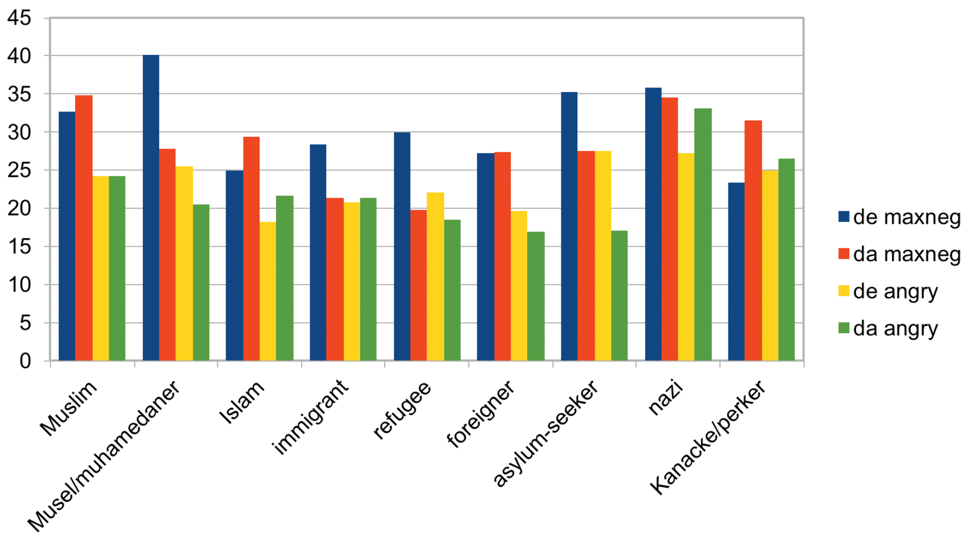

Fig. 1: Cross-language comparison of emoticon-negativity for key lexemes

languages, compounds like *Feminazi, Zionazi, Internazi* or even *grammarnazi* are frequent. In Danish, the word competes with the non-general Danish term *nazist,* which is the normal translation of German *Nazi* and has lower scores (21.5 for *angry* and 28.0 for maxneg).

Table 5 specifies the most related/specific emoticons for each lexeme, listing all scores above a threshold of 27.0, or - if no emoticon makes the threshold - the top scoring one. The most common top scorer was the *vomiting* emoji, but for the slur words *Alman* [de:German] and *perker* [da:foreigner], the insult emoji *middle-finger* outranked it. Another highly correlating emoticon class were animal emojis, with *pig-face* in particular being used for Muslims and asylum-seekers in Danish, while German has *camel* and *ogre* for immigrants and asylum-seeker, respectively. *Rat* and *monkey* appear above the 0.27 threshold for the Danish slurs *perker* and *nazi*.

Table 5: Keyword-emoticon closeness (p > 0.27, p < 0.27 in parentheses)

	GERMAN	DANISH
Moslem / muslim	vomiting 32.7, nauseated 29.2 (*Muslim*: vomiting 20.4)	pig-face 34.9, vomiting 27.8, skeptical 27.5, monkey 27.3
Musel / muhamedaner	vomiting 40.2, nauseated 34.1, middle-finger 33.3, camel 30.7, goat 30.7, ogre 30.3, pig-face 29.8	pig-face 27.8, vomiting 27.5
Islamist / islamist	vomiting 28.2	vomiting 29.2
Islam / islam	(vomiting 25.0)	vomiting 29.4
Migrant / immigrant	*Migrant*: vomiting 28.4 (*Immigrant:* thumbs-down 25.6)	(*immigrant*: skeptical 24.6) (*migrant*: angry 20.3)
Einwanderer / indvandrer	(camel 21.3)	(vomiting 19.1)
Flüchtling / flygtning	vomiting 30.0, thumbs-down 28.1	(vomiting 19.8)
Asylant / asylansøger	vomiting 35.2, thumbs-down 31.7, ogre 31.4, nauseated 30.8	(pig-face 20.2)
Ausländer / udlænding	(vomiting 25.6)	vomiting 27.4
Nazi / nazist	vomiting 35.9, middle-finger 33.9, nauseated 29.1, pig-face 27.2, angry 27.2	*nazist*: vomiting 28.0 *nazi*:vomiting 34.5, middle-finger 31.4, angry 33.1, pig-face 30.2, sad 30.6, rat 29.4, happy 28.6, monkey 27.3
Deutscher / dansker	thumbs-down 28.6, vomiting 28.5	thumbs-up 31.6, pig-face 30.1, happy 29.6, laugh 28.1
Alman / (nydansker)	middle-finger 34.6, angry 28.4, pig-face 28.1	(skeptical 14.2)
Kanacke / perker	(oncoming fist 26.0)	middle-finger 31.5, pig-face 31.1, vomiting 30.8, monkey 27.8, goat 27.6, rat 27.2

The radar chart in Fig. 2 provides an overview of word-embedding correlations between 15 common emoticons and the words *Muslim* and *refugee* in both languages. The most salient sentiments are *angry* and *vomiting*[14], as well as *pig-face* for Danish, and *thumbs-down* and *ogre* for German. Interestingly, Danish *muslim* (the red diamonds) exhibits stronger correlations, for both negative and positive emoticons, than Danish *flygtning* [refugee] (the green triangles). With a few exceptions (*vomiting*, *thumbs-down* and *ogre*), it also scores higher than the German counterparts. Also, *refugee* scores correlate with *Muslim* scores much more for German than for Danish, with negative sentiments for Danish *flygtning* being clearly less pronounced than for both its German counterpart *(Flüchtling)* and Danish *muslim*. Other findings from Fig. 2 are

14 Use of the *vomiting* emoji might be language-dependent. Both German and Danish, but not English, have a common adverbial pejorative construction, *zum Kotzen* (de) and *til at brække sig over* (da) [making you puke]. In addition, German uses *Kotz-* [puke-] as a pejorative prefix.

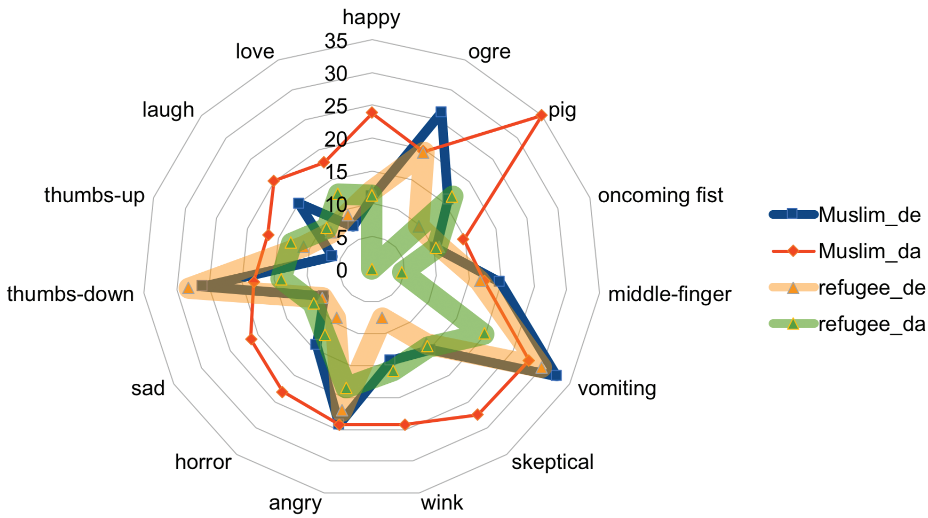

Fig. 2: Muslims and refugees, cross-language overview of emoticon sentiment

that the less emotional and more evaluative *thumbs-up* and *thumbs-down* show a much more marked difference for German than for Danish, and that pejorative animal emojis appear to be a bit language-specific ('pig-face' for Danish and 'ogre' for German). Finally, the topic of *refugee* seems not to be a laughing matter in either language, or at least less so than does *Muslim*. And where humour does come into play, there is a Danish lead for either term, in terms of 'laugh' and 'wink' emoticons.

Table 5 and Fig. 2 demonstrate that emoticon classes can paint a differentiated sentiment picture and that more than one class needs to be taken into account to assess overall negativity. But while animal emojis are dehumanizing and seem to function as pictorial 'slurs', no matter which animal is used, and while the strong negative sentiments *anger, vomiting, middle-finger* and *oncoming fist* nicely align with HS content, some negative emoticons are less interchangeable when it comes to assessing negativity. Thus, *laugh* can either be a happy, carefree signal, or a contemptuous reinforcer of a negative statement. Similarly, *wink* indicates irony, jest or an uncertain truth-value, but does not safely tell if the commented-on statement is positive or negative. Finally, *skeptical* is a negativity marker, but at another level than animal emojis, anger and aggressiveness, as it lacks their emotional force and thus affords an interesting angle on its HS assessment; for instance, it could be interpreted as a politeness indicator or a 'HS hedge' (*I'm not a racist, but ...*).

In order to shed light on the difference between skepticism and emotional negativity, Fig. 3 pits the *skeptical* emoticon class (x-axis) against the *vomiting* emoticon (y-axis) for comparable German and Danish terms from the immigration discourse. Where the German and English terms are mere translations of each other, English is used as a label (e.g *asylum-seeker* or *foreigner*); alternatively, if the English translation is ambiguous, the original words are used (e.g. *Kanacke/ perker* [muslim_slur]). In Fig. 3, an asterisk marks terms without a direct equivalent.

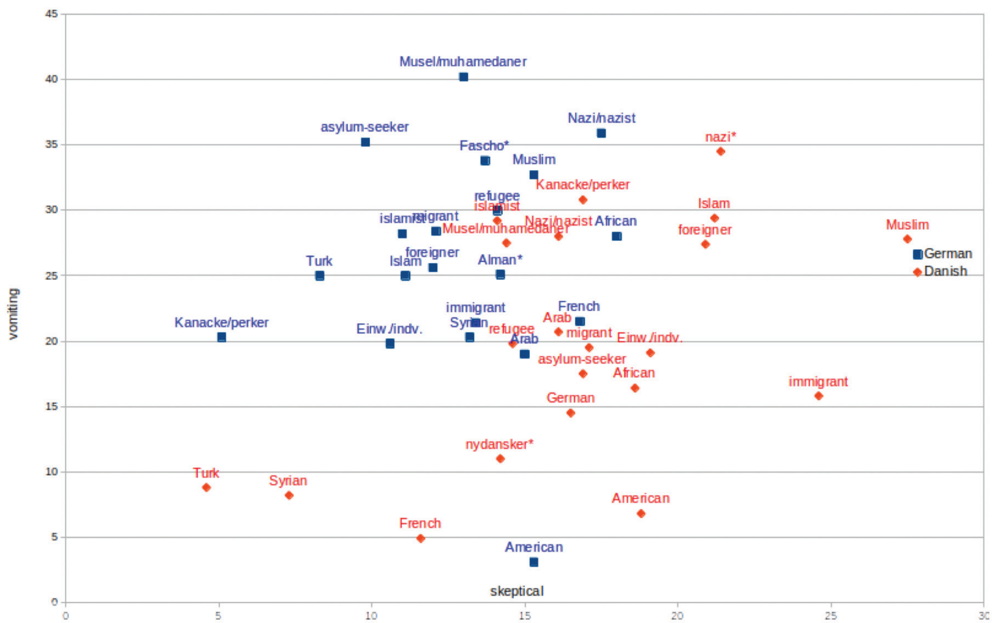

Fig. 3: Pejorative or skeptical or both? A German-Danish comparison

From the graphical representation in Fig. 3, it is immediately obvious that the word cloud of Danish is situated ‘south-east’ of the German one, suggesting that (on average!) equivalent terms are associated with less pejoration and more skepticism in Danish. The figure also shows the difference in where immigrants come from; thus, only *Arab* scores (slightly) worse in Danish than in German. *Turk* and *Syrian*, on the other hand, score irrelevantly low in the Danish data, but are situated near *Arab* in German, and close to *immigrant* as such. In both languages, *American* is viewed with moderate skepticism, but without negative emotion.

As mentioned in the discussion of Fig. 2 and Table 5, religion-linked immigration keywords *(Muslim, Islam, Islamist)* stand out as more emotionally negative in the Danish data, both compared to their German equivalents and to ethnic keywords or the word *refugee*. Fig. 3 shows that the difference is even bigger in terms of skepticism, especially for *Muslim*. By itself, immigration (such as in *immigrant, migrant* and in particular *asylum*-seeker) provokes more negative emotions in German, whereas in Danish it is still more linked to the *skeptical* emoticon.

Real slurs are difficult to match across languages, as the big difference between *Kanacke* and *perker* [foreigner_slur] indicates, with the former apparently being more harmless than the latter, in terms of emoticons. It is also interesting that while German *Musel* [muslim_slur] tops the *vomiting* scale, the 'equivalent' Danish *muhamedaner* is almost on par with the non-slur *Muslim* on that axis, and *less* subject to emoticon-skepticism than the latter. Possibly, the fact that *muhamedaner,* once the Danish word for *Muslim,* is marked as special is due to its low frequency and not being the currently sanctioned term, rather than by strong emotions. A slur that works in both languages is *nazi*. In Danish, this is the topscorer on the *vomiting* axis – a pure slur, since in non-slur uses the word is replaced with *nazist*. Finally, the neologism *nydansker* [new Dane], with its low *vomiting* score, is an example of an intentional non-slur, meant to be used instead of terms that have already acquired negative connotations.

7 Tagging evaluation

With its non-standard lexicon, orthography and syntax, CMC data are notoriously difficult to annotate. For German, this has been confirmed by several studies, e.g. Giesbrecht & Evert's (2009) 5-parser comparison and more recently Proisl (2018), both reporting around 93% POS accuracy for CMC with cross-domain training data. Even with web domain training data, Neunerdt (2013), while achieving 5% better POS accuracy than with a standard treebank, reported accuracy drops for chat and YouTube comments by 5% and 10%, respectively. It is relevant for our own research that CMC tagging results do not seem to be evenly distributed across word classes. Thus, also for German, Proisl found considerable performance drops for e.g. imperatives (80%), proper nouns (17.4%), and in particular, emoticons, half of which were not recognized. For Danish, no comparable studies were found.

For each language, a randomized sample (roughly 100,000 tokens per language) with immigrant/minority-related emoglyphs was evaluated for recall (R) and precision (P,). The former is calculated as the ratio of correctly recovered labels, the latter as the proportion of assigned labels that was correct. In addition, F1-measures are provided to sum up the two[15]. See Table 5.

Table 5: Precision/recall of emoglyph annotation

	PRECISION (P)	RECALL (R)	F1-MEASURE
German, emoticons	94.4	89.5	91.9
German, emojis	100	99.5	99.7
German, all	99.8	98.8	99.3
Danish, emoticons	91.7	95.7	93.7
Danish, emojis	100	97.9	98.9
Danish, all	99.5	96.6	98.0

15 F1 = 2 * R * P / (R + P)

Since emoji recognition was based on Unicode blocks, spurious readings were not likely (P=100). However, false negatives were still encountered where emojis outside these ranges were wrongly delimited and ended up as (foreign) ”letters” glued to words (e.g. ’☚In’)[16]. The distinction between punctuation and text emoticons was a bit more problematic (F=91.9), with an accuracy more similar to the average reported for ordinary POS in CMC data. For instance, the unlimited addition of extra parentheses for emphasis was not foreseen in the evaluated version of the tagger, so that ’;))))’ was read as ’;))’ and two individual punctuation parentheses, causing both a false positive and a false negative error at the same time. Note that this particular error is difficult to solve completely - after all, this could have been an emoticon at the end of a parenthesis enclosure.

8 Conclusions and outlook

We have shown how a fine-grained sentiment annotation of emoticons/emojis can be integrated into the linguistic mark-up of a large bilingual Social Media corpus with a reasonable tagging accuracy and used to facilitate the inspection and identification of hate speech usage, as well as quantify cross-language differences and minority-triggered sentiments. Not least for non-direct hate speech, sarcasm and irony, emoticon annotation adds a valuable new dimension to corpus searches and excerpting.

While the current annotation primarily aimed at sentiment-marking and truth-value-flagging of statements or entire tweets, future work needs to address emoji syntax at a more detailed level, ascertaining whether a given symbol is linked to the whole clause or to a specific word, or whether it is functioning as a word itself. Also, in the face of a growing use of multi-emoji strings, complex emoji meaning should be explored and list usage (co-ordinated usage) distinguished from inter-emoji modifier relations.

References

Abadi, Martin et al. 2016. TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. 265-283.

Basile Valerio, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco M. R. Pardo, Paolo Rosso& Manuela Sanguinetti. 2019. SemEval-2019 Task 5: Multilingual Detection of Hate Speech Against Immigrants and Women in Twitter. In Jonathan May et al. (Eds.) *Proceedings of the 13th Int. Workshop on Semantic Evaluation (SemEval-2019), co-located with the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), Minneapolis, Minn.,, June 6-7*. 54-63.

Baumgarten, Nicole, Eckhard Bick, Klaus Geyer, Ditte A. Iversen, Andrea Kleene, Anne Vibeke Lindø, Jana Neitsch, Oliver Niebuhr Rasmus Nielsen & Esben N. Petersen. 2019. Towards

16 Incidentally, the resulting ”word” could not be analyzed either, causing collateral errors in addition to the emoji problem.

Balance and Boundaries in Public Discourse: Expressing and Perceiving Online Hate Speech (XPEROHS). In Jacob L. Mey, Alexandra Holsting & Christian M. Johannessen (Eds.) *RASK - International Journal of Language and Communicationvol.* 50. University of Southern Denmark. 87-108.

Beißwenger, Michael, Sabine Bartsch, Stefan Evert & Kay-Michael Würzner. 2016. EmpiriST 2015: A shared task on the automatic linguistic annotation of computer-mediated communication and web corpora. In Paul Cook et al. (Eds.) *Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task*. Berlin: Association for Computational Linguistics. 44-56.

Bick, Eckhard & Tino Didriksen. 2015. CG-3 – Beyond Classical Constraint Grammar. In Beáta Megyesi (Ed.): *Proceedings of NODALIDA 2015, May 11-13, 2015, Vilnius.*. Linköping: LiU Electronic Press. 31-39.

Bick, Eckhard. 2020. An Annotated Social Media Corpus for German. In Nicoletta Calzolari et al. (Eds.) *Proceedings of the 12th International Conference on Language Resources and Evaluation, LREC2020 (Marseille, May 2020)*. ACL / ELRA. 6129-6137.

Cappallo, Spencer, Stacey Svetlichnaya, Pierre Garrigues, Thomas Mensink & Cees G. M. Snoek. 2019. New Modality: Emoji Challenges in Prediction Anticipation and Retrieval. *IEEE Transactions on Multimedia vol. 21, no. 2*. 402-415.

Felbo, Bjarke, Alan Mislove, Anders Søgaard, Iyad Rahwan & Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting emotion, sentiment and sarcasm. In Martha Palmer et al. (Eds.) *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. ACL. 1615-1625.

Foxman, Abraham H. & Christopher Wolf. 2013. *Viral Hate: Containing Its Spread on the Internet.* New York: St. Martin's Press.

Giesbrecht, Eugenie & Stefan Evert. 2009. Is Part-of-Speech Tagging a Solved Task? An Evaluation of POS Taggers for the German Web as Corpus. In Iñaki Alegria et al. (Eds.) *Proceedings of the 5th Web as Corpus Workshop (WAC5), San Sebastian, Spain.*

Hauthal, Eva, Dirk Burghardt & Alexander Dunkel. 2019. Analyzing and visualizing emotional reactions expressed by emojis in location-based social media. *Journal of Geo-Information 8(3)*. 1–21.

Herz, Michael & Peter Molnar (Eds.). 2012. *The Content and Context of Hate Speech. Rethinking Regulation and Responses*. Cambridge: Cambridge University Press.

Kratzke, Nane. 2017. #BTW17 Twitter Data-set - Recorded Tweets of the Federal Election Campaigns of 2017 for the 19th German Bundestag. Data 2017, 2(4), 34 [Data set]. https://doi.org/10.3390/data2040034.

Ljubešic, Nikola & Darja Fišer. 2016. A global analysis of emoji usage. In Paul Cook et al. (Eds.) *Proceedings of 10th Web as Corpus Workshop*. Association for Computational Linguistics. 82-89.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, G. S. Corrado, & Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In C.J.C. Burges et al. (Eds.) *Advances in neural information processing systems - Proceedings of NIPS 2013*. 3111–3119.

Neunerdt, Melanie, Bianka Trevisan, Michael Reyer & Rudolf Mathar. 2013. Part-of-Speech Tagging for Social Media Texts. . (2013). Language Processing and Knowledge in the Web. pp. 139-150. Springer

Novak, Petra K., Jasmina Smailović, Borut Sluban & Igor Mozetic. 2015. Sentiment of Emojis. *PloS one, 10(12):e0144296*

Pereira-Kohatsu, Juan Carlos, Lara Quijano-Sánchez, Federico Liberatore & Miguel Camacho-Collados. 2019. Detecting and Monitoring Hate Speech in Twitter. *Sensors (Basel, Switzerland), 19(21), 4654*.

Proisl, Thomas. 2018. SoMeWeTa: A Part-of-Speech Tagger for German Social Media and Web Texts. In Nicoletta Calzolari et al. (Eds.) *Proceedings of LREC 2018*. 665-670.

Wang, Hao & Jorge A. Castanon. 2015. Sentiment expression via emoticons on social media. In Howard Ho et al. (Eds.) *Proceedings of IEEE International Conference on Big Data (Santa Clara)*. 2404-2408.

Waseem, Zeerak & Dirk Hovy. 2016. Hateful Symbols or Hateful People? - Predictive Features for Hate Speech Detection on Twitter. In *Proceedings of the NAACL Student Research Workshop (San Diego, California)*. ACL. 88-93.

Wiegand, Michael, Melanie Siegel & Josef Ruppenhofer. 2018. Overview of the GermEval 2018 Shared Task on the Identification of Offensive Language. In Josef Ruppenhofer et al. (Eds.) *Proceedings of GermEval 2018, 14th Conference on Natural Language Processing (KONVENS 2018), Vienna*. 1-10.