

Emoticons as multifunctional and pragmatic Resources: a corpus-based Study on Twitter

Stefania Spina

University for Foreigners of Perugia
Piazza Fortebraccio, 4 – Perugia, Italy
E-mail: stefania.spina@unistrapg.it

Abstract

Emoticons play an important role in digital written communication: they can serve as markers either of emotions or social relationship and familiarity, and they can intensify or downgrade the pragmatic force of a text.

The aim of this study is to investigate the use of emoticons in Twitter by Italian users, and to verify, by relying on corpus data and on statistical methodologies, some of the prevailing opinions on the use of emoticons: that they are technically-driven resources, that they are mostly used by young people, and more often by females, and that they are superficial and easy ways of expressing emotions using images instead of words.

A mixed-effects model analysis has shown that the use of emoticons on Twitter is affected by a complex interaction of cultural, technological, situational and sociolinguistic variables.

Keywords: emoticons, Twitter, mixed-effects models.

1. Introduction

Emoticons (the graphic signs, such as the smiley face, that often accompany digital written communication) are an integral part of digital culture since its beginnings: they have followed its development over the last decades, evolving alongside with the rapid spread of new written communication environments, such as social media or text messaging systems.

Many studies have outlined the key role of emoticons in digital written communication (e.g. Amaglobeli, 2012; Baron, 2009; Danesi, 2016; Derks, Bos & von Grumbkow, 2007; Dresner & Herring, 2010; Spina, 2016; Vandergriff, 2014; Walther & D’Addario, 2001; Yus & Yus, 2014): as people use writing more and more instead of face to face interactions or phone calls, the need for overcoming limitations in communicating emotional tone arises. The widespread use of emoticons allows to convey nonlinguistic information that in face-to-face communication is expressed through facial expression and other bodily indicators. Emoticons, therefore, are primarily “emotion icons”: additional opportunities to convey emotions through the use of graphic symbols, directly mapped onto facial expressions.

The role of emoticon in digital written communication, however, is much more nuanced and not limited to the expression of emotions. Following Dresner & Herring (2010), Vandergriff (2014), and Spina (2016), they are developing at least two other important pragmatic functions, that are not necessarily mapped onto facial expressions, or aimed at the expression of emotions:

- the function of social markers of familiarity and empathy. In this sense, they are relational icons, that promote rapport and play a social and affiliative role;

- the function of markers of the pragmatic force of a text, aimed at intensifying or downgrading its meaning. In this function, they are contextualization cues (Gumperz, 1982; Auer, 1992), that provide information on how to interpret the verbal message.

As a consequence, emoticons are multifunctional and highly context-sensitive resources, whose different functions most often tend to overlap and to occur simultaneously within the use of a single emoticon. This claim is illustrated by the examples (1) and (2):

(1)
@user2 Ci vieni in piscina domani?
[@user2 Are you coming to the pool tomorrow?]

@user1 No :-(

(2)
@user2 Hai visto la foto del mio profilo?
[@user2 Have you seen my profile picture?]

@user1 Bellissima!! :-)))
[@user1 So beautiful!! :-)))]

In example (1), the sad emoticon serves both as a mitigation resource, aimed at softening the refusal of an invitation, and as a means of expressing regret for this refusal. In example (2), the smiley is both a marker of intensification of the positive emotion expressed verbally by “so beautiful” and graphically by the exclamation marks, and a marker of familiarity, aimed at expressing empathy and friendliness. Emoticons, therefore, are not just a ludic and extralinguistic supplement to language, with the exclusive role of expressing emotions, but rather linguistic resources that

play other important pragmatic functions in digital written communication, such as conveying the intentions of the writer (Tagg, 2012), supporting social relationships among participants, and providing new opportunities for creative expressions.

2. Motivation

The main aim of the present investigation was to examine the use of emoticons in Twitter by Italian users. More specifically, it tried to verify, by relying on corpus data and on statistical methodologies, some of the prevailing findings on the use of emoticons in digital written communication: that they are technically-driven resources, whose spread is mainly due to the diffusion of mobile devices (Baron, 2008); that they are mostly used by young people (Merchant, 2001; Tagliamonte & Denis, 2008), and more often by females (Baron, 2008; Huffaker & Calvert, 2005; Spina & Cancila, 2013; Tossel et al., 2012); finally, that they are easy ways of expressing emotions using images instead of words (Provine, Spencer & Mandell, 2007).

Conversely, the nuanced social and pragmatic functions played by emoticons in digital communication suggest that their use and distribution should be affected by a more complex interaction of technological, cultural, situational and sociolinguistic variables.

The questions that this study tried to answer were: what are the variables that, at a discourse level, affect the use of emoticons in Twitter interactions? How is the use of emoticons influenced by these variables?

3. Method

To answer these questions, a large corpus of tweets extracted from the Italian timeline was used. The Ita_twitter corpus (Spina, 2016) contains more than 550,000 tweets sent in a time span of seven months (November 2012-May 2013). The 8,842,450 tokens were pos-tagged through an ad hoc version of TreeTagger (Schmid, 1994), purposely trained to automatically detect emoticons.

From the Ita_twitter corpus, a subset written by 290 users was randomly selected. This subset consists of 4,441 tweets and contains information on the authors (sex, geographical provenance), on their level of mastery within Twitter environment (date of registration on Twitter, number of tweets sent), on the technical context (the software device from which each tweet was sent), and on the type of tweet (a simple status update, or the reply to a previous tweet written by someone else). Information on the authors' age was obtained by manually checking each of the 290 profiles. The subset of 4,441 tweets contained 15 different types of ASCII emoticons, that are listed in table 1. Each of the types is represented in the corpus by a number of different graphic forms, depending on the combination of ASCII symbols. The classic smiley :-), for example, is represented by a number of different forms, such as :), :)), :-)), etc.

Emoticon	Meaning
:-)	smile
;-)	wink
:-(sad or frown
:*	kiss
<3	heart
_	dazed
:')	tears of happiness
^_^	happy
:P	tongue sticking out
x.x	dead
:'(crying
-.-	annoyed
:D	laughing
O.O	surprised
u.u	sarcastic

Table 1: the 15 types of emoticons used in the corpus

In addition, given that the corpus was pos-tagged and lemmatized, a range of other linguistic information could be added to the selected tweets, including the type of sentence (question, exclamation, etc.), the co-occurrence of other discourse elements relevant to Twitter interactions (hashtags, mentions), and the length of each tweet (in number of tokens).

In order to explore how Italian participants use emoticons in their Twitter interactions, a mixed-effects model analysis was performed on the selected data. Mixed-effects modeling (e.g., Baayen, Davidson, & Bates, 2008) is particularly suited to corpus data (Gries, 2015), because it can integrate multiple categorical and numeric variables (fixed effects), and, at the same time, it can address the idiosyncrasies deriving from the analysis of data produced by the same subjects (random effects).

The mixed-effect model was built using R version 3.3.3 and the R packages lme4 (version 1.1-13; Bates, Maechler, & Bolker, 2012), lmerTest (version 2.0-33), and sjPlot (version 2.3.1). The number of emoticons used in each tweet (range: 0-27, mean: 0.26, sd 0.68) was used as dependent variable, and the following predictors were initially included in the model: the age (range: 16-67, mean: 31.72, sd 10.57) and sex of participants (f: 1443; m: 2998); the device from which the tweets were sent (mobile or desktop); the level of mastery within the Twitter environment (measured as the number of tweets sent from the date of registration to the date of each tweet in the corpus), distributed in five bands, from the lowest (a) to the highest (e); the type of tweet (status update: 1842, or reply: 2599); the number of co-occurring hashtags (range: 0-8, mean: 0.23; sd 0.63); the tweet length in number of tokens

(range: 1-41, mean: 13.85, sd 7.16), and the type of sentence (declarative: 3053, or non-declarative, that is exclamative or interrogative: 1388).

In order to model the individual differences in the use of emoticons, the authors of the tweets were used as random effect, by assuming different random intercepts for each author. In addition, as random slope models allow the predictors to have a different effect for each subject, random slopes were included in the model, with the aim of accounting for the different effects that the “type of tweet” and “sentence” variables have on each single author (Winter, 2013).

Finally, in the model building process, a backward selection approach was adopted, starting with a full model, including all the fixed effects mentioned above, and then dropping one variable at a time, and excluding a variable from the model if non-significant (Gries, 2015).

4. Results

Preliminary results show that four predictors affect the use of emoticons in Twitter as fixed effects (see the plot in figure 1). The predictor with the stronger effect is the type of tweet: emoticons occur far more in replies to the tweets of other participants, which are the most interactive form of tweet (Honeycutt & Herring, 2009; Schnoebelen, 2012), than in status updates. This finding clearly confirms the hypothesis that emoticons are one of the linguistic resources that participants rely on when they need support for establishing or maintaining a social relationship with their interlocutors. Replies, that automatically include a mention to the addressee, are in fact one-to-one or one-to-few interactions, even if they also presuppose the presence of the more numerous audience of followers. In this context, emoticons seem particularly suited to convey reactions to opinions or feelings expressed by others.

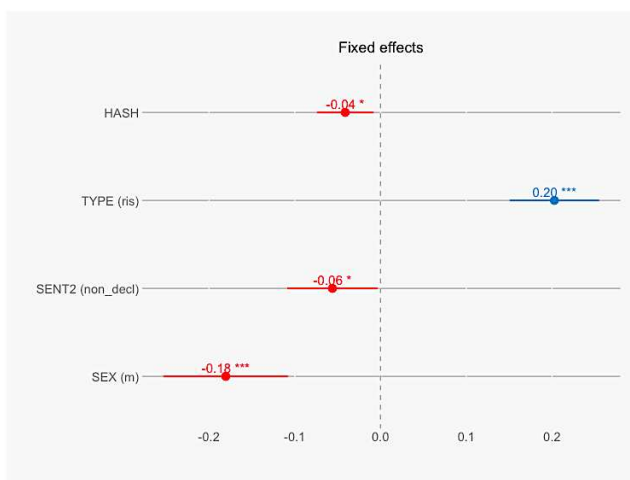


Figure 1: A plot of the estimates of the four significant fixed effects.

Another highly affecting predictor is sex: females use emoticons significantly more than males. This finding confirms previous research in computer-mediated communication (e.g. Vandergriff, 2014; Wolf, 2000; but

Danesi, 2016), according to which “females are the prime users of emoticons” (Baron, 2008:65). The same is not true, however, for the age of participants: this variable is never significant in the model as a fixed effect, and cannot therefore be considered a major predictor of the use of emoticons in Twitter.

The different types of sentence used in the tweets are another significant predictor; the model shows that the presence of exclamative or interrogative sentences, signalled by the ? and ! punctuation marks, produces a decrease in the use of emoticons. This finding could seem surprising, since emoticons are a graphical means of expressing emotions, and exclamatives often convey feelings and emotions, but it can be interpreted as an evidence of the multifunctionality of emoticons: one of their roles in digital written communication is that of syntactic markers, often serving as punctuation in place of traditional punctuation marks (Amaghlobeli, 2012). We can conclude, then, that exclamative and interrogative sentences decrease where emoticons replace question and exclamation marks, as in examples (3) and (4):

(3)
 @user ciao... Buon lunedì... Che bell'inizio di settimana :-D
 [@user Hello... Have a good Monday! What a great start of the week :-D]

(4)
 @user Ma che ci fai per DUE MESI a new york :D
 [@user What are you going to do for TWO MONTHS in New York :D]

The last significant fixed effect of the model is the hashtag. Emoticons and hashtags tend to have a complementary distribution: when more emoticons are used, less hashtags are found in the tweets. This finding seems coherent with the respective functions of the two discourse elements: while emoticons express either emotions or familiarity, or mark the pragmatic force of a text, the hashtag has an informative function (marking the topic of a text), or addresses the social need of aggregating communities of participants around a common theme or interest (Zappavigna, 2015). In this sense, emoticons seem to serve the pragmatic function of supporting social relationships among few participants, whereas the hashtag plays an important role in affiliating large masses of people in flows of conversations on shared topics.

Going further with the analysis, the picture described so far gets clearer if interactions between different predictors are considered. The type of tweet and the level of mastery, for example, have an interaction effect on the use of emoticons: while replies always contain a greater number of emoticons, this effect seems to slightly increase if the users are more familiar with Twitter. In the case of status updates, the opposite is true: the less proficient the users, the less emoticons they use. In addition, the mastery of Twitter rules and conventions also interacts with the length of tweets in the effect on the use of emoticons: as shown in Figure 2, low mastery levels (a) produce shorter tweets with less emoticons; medium and high mastery levels (b, c, d and e), conversely, tend to increase the number of

emoticons, together with the text length.

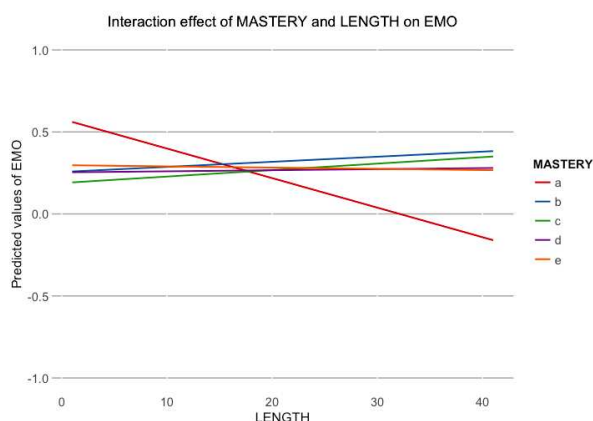


Figure 2: A plot of the interaction between mastery and length of tweets in their effect on the use of emoticons.

From the above, we can conclude that the use of emoticons in Twitter is not technically-driven (the predictor “software device” is never significant in the model, neither as a fixed effect nor in interaction with other factors), and it is not exclusive of young people; rather, it is influenced by a number of pragmatic, cultural and sociolinguistic factors interacting with each other. As a result, far from being only an add-on feature or a frivolous way of expressing emotions, emoticons are constitutive of CMC (Vandergriff, 2014), since they are assuming more and more sophisticated social and pragmatic roles.

After this quantitative investigation, a more in-depth and qualitative analysis needs to be conducted on emoticons, in order to investigate in more detail the linguistic context that favors their use, and their distribution among the previously mentioned functions. A specific attention should be paid to the linguistic features that are traditionally associated to the purpose of establishing and maintaining relationships with other participants (discourse markers and personal pronouns, for example) and of modulating the pragmatic force of texts (intensifiers, affective vocabulary, etc.).

5. References

- Amaghlobeli, N. (2012). Linguistic features of typographic emoticons in SMS discourse. *Theory and Practice in Language Studies*, 2(2), pp. 348--354.
- Auer, P. (1992). *The contextualization of language*. Amsterdam: Benjamins.
- Baayen, R. H., Davidson, D. J. and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, pp. 390--412.
- Bates, D. M., Maechler, M. and Bolker, B. (2012). *lme4: Linear mixed-effects models using Eigen and S4*.
- Baron, N. (2008). *Always On: Language in an Online and Mobile World*. Oxford: OUP.
- Baron, N. (2009). The Myth of Impoverished Signal: Dispelling the Spoken Language Fallacy for Emoticons in Online Communication. In J. Vincent & L. Fortunati (Eds.), *Electronic Emotion: The Mediation of Emotion via Information and Communication Technologies*. Bern: Peter Lang, pp. 107--135.
- Baron, N. (2010). *Alphabet to Email: How Written English Evolved and Where It's Heading*. London: Routledge.
- Danesi, M. (2016). *The Semiotics of Emoji. The Rise of Visual Language in the Age of the Internet*. London: Bloomsbury
- Derks, D.; Bos, A.E.R and von Grumbkow, J. (2007). Emoticons and social interaction on the Internet: the importance of social context. *Computers in Human Behavior*, 23, pp. 842--849.
- Dresner, E., Herring, S.C. (2010). Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force. *Communication Theory*, 20(3), pp. 249--268.
- Gries, S. Th. (2015). The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10(1), 95--125.
- Gumperz, J. (1982). *Discourse strategies*. Cambridge: CUP.
- Honeycutt, C., Herring S.C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. In *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)*, Los Alamitos: IEEE Press.
- Huffaker, D. A., and Calvert, S. L. (2005). Gender, Identity, and Language Use in Teenage Blogs. *Journal of Computer - Mediated Communication*, 10(2).
- Lo, S.K. (2008). The Nonverbal Communication Functions of Emoticons in Computer-Mediated Communication. *CyberPsychology & Behavior*, 11(5), pp. 595--597.
- Merchant, G. (2001). Teenagers in cyberspace: An investigation of language use and language change in Internet chatrooms. *Journal of Research in Reading*, 24(3), 293--306.
- Provine, R.R.; Spencer, R. and Mandell, D. (2007). Emotional expression online: Emoticons punctuate website text messages. *Journal of Language and Social Psychology*, 26(3), pp. 299--307.
- Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*.
- Schnobelen, T.J. (2012). Emotions Are Relational: Positioning and the Use of Affective Linguistic Resources. Ph.D. thesis, Department of Linguistics, Stanford University.
- Spina, S. (2016). *Fiumi di parole. Discorso e grammatica delle interazioni scritte in Twitter*. Loreto: StreetLib.
- Spina, S., Cancila, J. (2013). Gender issues in the interactions of Italian politicians on Twitter: Identity, representation and flows of conversation. *International Journal of Cross-cultural Studies and Environmental Communication*, 2(2), pp. 147--157.
- Tagg, C. (2012). *Discourse of Text Messaging: Analysis of SMS Communication*. London: Continuum.
- Tagliamonte, S., Denis, D. (2008). Linguistic ruin? LOL! Instant Messaging and teen language. *American Speech*, 83(1), pp. 3--34.
- Tossell, C. C., Kortum, P., Shepard, C., Barg-Walkow, L. H., Rahmati, A., and Zhong, L. (2012). A longitudinal

- study of emoticon use in text messaging from smartphones. *Computers in Human Behavior*, 28(2), pp. 659--663.
- Vandergriff, I. (2014). A pragmatic investigation of emoticon use in nonnative/native speaker text chat. *Language@ Internet*, 11.
- Walther, J., D'Addario, K. (2001). The impacts of emoticons on message interpretation in computer mediated communication. *Social Science Computer Review*, 19, 324--347.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [<http://arxiv.org/pdf/1308.5499.pdf>].
- Wolf, A. (2000). Emotional Expression Online: Gender Differences in Emoticon Use. *CyberPsychology & Behavior*, 3(5), pp. 827--833.
- Yus, F., and Yus, F. (2014). Not all emoticons are created equal. *Linguagem Em (Dis) Curso*, 14(3), pp. 511--529
- Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, 25(3), pp. 274--291.