Signaling irony, displaying politeness, replacing words

The eight linguistic functions of emoji in computer-mediated discourse

Louis Escouflaire
Université Catholique de Louvain

This paper aims to identify and define the different linguistic functions that emoji can carry in French and English computer-mediated discourse (CMD). Emoji are graphical icons, representing attitudes or concepts, that can be used on any smartphone or computer with a variety of communicative roles inside a conversation. Using a multimodal corpus containing 1200 emoji and the messages in which they were used, we tested an initial typology of emoji functions inspired by previous frameworks. The quantitative analysis of the data, along with the detailed examination of specific examples, led to the construction of a new, more refined typology consisting of three primary functions (expressive, interpretative and referential) and five secondary functions (relational, politeness, emphatic, structural and aesthetic).

Keywords: emoji, conversation, social media, computer-mediated communication, discourse analysis

1. Introduction

In 1982, computer scientist Scott Fahlman realized that computer-mediated communication (CMC; Herring 1996) lacked something important: a way to signal whether a written message is intended as ironic or not. In face-to-face conversation, paraverbal cues such as intonation, laughter or facial expressions can help the receiver understand when the writer is being sarcastic (Troiano & Nante 2018). As these cues are not available in CMC, Fahlman proposed to add the string of characters ":-)" to a message in order to indicate that this message has to be interpreted ironically (Dresner & Herring 2010). Emoticons were born.

https://doi.org/10.1075/li.00062.esc
Lingvisticæ Investigationes 44:2 (2021), pp. 204–235. ISSN 0378-4169 | E-ISSN 1569-9927

These smiley faces evolved into "emoji" (from Japanese "e" 絵, "picture"; and "moji" 文字, "character")), which were incorporated in Apple's iPhone in Japan in 2007 and in Europe in 2011. Emoji can be defined as graphical icons, representing attitudes or concepts, that can be used in computer-mediated discourse (CMD; Herring 2001) in order to fill a variety of communicative roles inside a message or a conversation. Since their inclusion in the Unicode Standard in 2010, 3.304 different types of emoji are used every day by millions of CMD users all over the world.[1] In 2015, the Oxford Dictionaries elected the "face with tears of joy" emoji ("😂") as their "word of the year", arguing that "emoji have come to embody a core aspect of living in a digital world that is visually driven, emotionally expressive, and obsessively immediate" (Steinmetz 2015).

In this paper, we analyze the different roles that emoji can play in CMC in order to build a typology of the linguistic functions of emoji. Even though emoticons were initially created as disambiguation tools for ironic messages, it has been demonstrated that they are able to perform a wide variety of conversational functions that users consciously or unconsciously apply and understand (Herring & Dainas 2017; Spina 2018; Beißwenger & Pappert 2019). However, no academic consensus has yet been found on a classification scheme for emoji functions. The goal of this paper is to build a new fine-grained typology of the linguistic functions of emoji and to evaluate its relevance on a CMC corpus of French and English social media messages.

Section 2 of this paper aims to provide a synthetic compilation of the most prominent typologies proposed for functional analysis of emoticons and emoji in CMD. The preliminary framework obtained through this review of the literature contains seven different functions, which are described one by one in detail with examples in the same section. This synthetic typology is then applied to a multimodal corpus containing 1200 emoji used on Facebook and on Twitter in French and English messages, as described in Section 3. Section 4 of this paper is devoted to the presentation of our findings, leading to the description of our final typology of emoji functions. In concluding, we address some limitations of our study and propose perspectives for further research.

2. The linguistic functions of emoji

This section presents the seven linguistic functions of emoji in CMD that were identified by comparing a variety of papers produced on emoticons and emoji since 2007. It was decided to include some studies investigating emoticons (e.g.,

1. https://emojipedia.org/stats/, last accessed on January 15, 2021.

":-)") among those focusing on emoji (e.g., "😊"), as Herring & Dainas (2017) state that emoji are able to fulfill the same conversational roles as emoticons (and more). Although the visual and practical differences between emoticons and emoji could lead to slight functional disparities, this has not been observed by any researcher. Still, it should be noted that emoticons predominantly represent human faces and hearts, while emoji can be used to represent a much wider range of items and concepts. This is shown by the fact that eight different "emoji categories" are used on most digital keyboards: "Smileys & People", "Animals & Nature", "Food & Drink", "Activity", "Travel & Places", "Objects", "Symbols" and "Flags". However, "Smileys & People" emoji are the most commonly used emoji in conversation (Unicode 2019), and those that are able to display the most different linguistic functions. The other seven emoji categories are almost exclusively used in a referential way, as explained in Section 2.7.

Some of the papers which inspired the preliminary typology presented in this section identified only three or four different emoji functions (Marcoccia & Gauducheau 2007; Cramer et al. 2016; Dürscheid & Siever 2017; Tang & Hew 2019), while others distinguished up to eight or nine functions (Herring & Dainas 2017; Beißwenger & Pappert 2019). Panckhurst & Frontini (2020) use a three-fold classification based on the ways emojis are used to construct the relationship between the scriptor and the recipient of the message. Na'aman et al. (2017) suggested that emoji can have two major functions, with some sub-functions included in the framework. Our preliminary typology aims to constitute a common ground between the different studies that influenced it. The seven functions that constitute it are presented one by one in this section, along with examples of their use in CMD messages.

The examples used in this section were made up and inspired by messages found in the corpus presented in Section 3.1.

2.1 The expressive function

One of the most natural and obvious functions of emoji (and emoticons) relies on their ability to bring CMD closer to oral communication by mimicking paraverbal cues, such as facial expressions and gestures. Expressive emoji allow writers to convey emotions in a conversation or to add emotional value to verbal content. These emotions are mapped directly onto physical expression (Dresner & Herring 2010): happiness can be expressed with a "grinning face" emoji ("😀"), anger with an "angry face" emoji ("😠"), or support with a "raised fist" emoji ("✊"), for example (Schneebeli 2017).

(1) *I'm going to a party tonight* 😁

(2) a. *What did you think of the main character's death?*
b. 😭😭😭

There are two different ways in which emotions can be expressed through the use of emoji in CMD. First, expressive emoji are able to indicate the emotional state of the writer when sending the message (Marcoccia & Gauducheau 2007). In this case, they work as "indicators of affective states" (Dresner & Herring 2010; Na'aman et al. 2017), as shown in Example 1 above. Then, emoji can also expose the emotional value of a message when no emotion is present in the verbal part of the text (Cramer et al. 2016). In this case, emoji play the role of "indicators of stance", showing how the writer stands in the conversation (Schneebeli 2017). In the same way, they can be used to react to a previous message in the conversation, or to a situation, as in Example 2. In Facebook comment threads, for instance, emoji are often found on their own (without verbal content) for reacting to a previous message in the thread (Herring & Dainas 2017). Beißwenger & Pappert (2019) call this use the "evaluative" or "commentary" function of emoji. Panckhurst & Frontini (2020) differentiate between emoji used to convey the scriptor's own feelings and emoji referring to the recipient's feelings.

Quite logically, "Smileys and People" emoji (in particular those representing facial expressions and attitudes) are the main types of emoji to which the expressive function can be assigned, as they represent iconically how the writer's emotions would appear in a face-to-face conversation. However, some non-facial emoji can also be indicators of emotional state or stance. Gesture emoji, such as the "thumbs up" ("👍") or the "clapping hands" emoji ("👏"), can display agreement and positive reaction from the writer, while the "folded hands" emoji ("🙏") can express admiration or gratitude. The "red heart" emoji ("❤️") (along with its many variants) is also used by CMD writers to express love or simply a strong positive feeling towards something or someone (Schneebeli 2017).

2.2 The interpretative function

In face-to-face conversation, non-verbal cues like facial expressions and laughter are not only tools for expressing emotions, but also for altering the meaning of a message. In CMD, emoji can be used to replace these features and work as indications to how a sent message should be understood by the receiver (Herring & Dainas 2017). This relates historically to the initial purpose of emoticons (i.e., disambiguating ironical messages), as attributed by their creator Scott Fahlman. Some researchers state that interpretative emoji act as "tone" modifiers (Cramer et al. 2016), in opposition to expressive emoji which affect the "mood" of the mes-

sage. Dürscheid & Siever (2017) refer to this particular use of emoji as "intention signals".

(3) *Mom said we're eating pizza again today* 😁

(4) *Mom said we're eating pizza again today* 😪

These emoji guide the interpretation of a message and help clarify the intention of the writer and eliminate potential misunderstandings between interlocutors (Amaghlobeli 2012). Sentences 3 and 4 above show how the presence of an emoji in a message can influence its interpretation: the "grinning face with smiling eyes" emoji in Example 3 suggests that the writer is excited about the situation, whereas the same sentence in Example 4, with the "crying face" emoji at the end, rather guides the receiver towards a negative interpretation of the message. Spina (2018) refers to such emoji as pragmatic markers, as they allow users to "infer contextually appropriate meanings in text."

A context in which emoji are often used with their interpretative function is that of playfulness, such as irony or sarcasm. In CMD, emoji can be used as effective markers of irony, in order to let the receiver understand that the intended meaning of the message is different from its verbal content. Research has shown that emoji are indeed the most efficient way to identify and decode sarcastic messages in CMD (González-Ibáñez et al. 2011; Weissman & Tanner 2018).

(5) *I hate you so much* 😂

Dresner & Herring (2010) introduced an interesting approach to emoticons, which was later applied to emoji by Schneebeli (2017). Inspired by the speech act theory developed by Austin (1962) and Searle (1969), they analyzed emoji as "illocutionary force markers", indicators of the intended action of a message: interpretative emoji "help convey the speech act performed through the production of the utterance" (Dresner & Herring 2010). Following this approach, emoji can be considered as a way to make the implicit more obvious (Schneebeli 2017). In sentence 5, for instance, the "face with tears of joy" emoji indicates that the statement should not be taken literally, because the writer is being ironic.

2.3 The relational function

Emoji can also be used to maintain and manage relationships between CMD writers. Like eye-contact or laughter in face-to-face conversation, they can play a phatic role and work as indicators of closeness between users (Marcoccia & Gauducheau 2007), or markers of familiarity (Spina 2018). Messages accompanied by emoji are perceived as more pleasant and have been shown to display a higher level of intimacy (Janssen et al. 2014; Tang & Hew 2019).

Furthermore, they can work as ways to request familiarity from receivers, which they can accept by using emoji in their answer as well. Moreover, in the same way as small talk and ritualized expressions like "hello" or "how are you?", the use of emoji participates in the construction of a positive image of the writer (Spina 2018). Emoji users tend to be considered as "more extravert and agreeable" than those who do not include emoji in their messages (Fullwood et al. 2015).

(6) a. *The exam is at 10 am tomorrow*
b. *Thanks!!*
a. *No problem* 😊
b. 😊

Emoji can also have an "economizing" phatic function, maintaining a "conversational connection" between interlocutors when no words are left to be said (Beißwenger & Pappert 2019). For example, sending a single emoji to someone can be interpreted as a way of showing them that you are available for conversation, or simply that you are thinking about them (Kelly & Watts 2015). Similarly, emoji are often used on their own in order to end a conversation, as a final conversational turn. They signal that the previous message was acknowledged by the writer and that he or she wants to put an end to the conversation: "when not knowing what to say or not wanting to say anything, or when saying nothing would be inappropriate" (Cramer et al. 2016). Example 6 above illustrates such a conversation: *b* ends the conversation with a single relational emoji, acknowledging *a*'s second turn. *B* also appears to accept *a*'s request for familiarity by using the same emoji that was used by *a*. Aull (2019) refers to such emoji as "phatic tokens".

2.4 The politeness function

Another function of emoji relates to their capacity to work as linguistic hedges, according to the politeness theory introduced by Brown and Levinson (1987), who analyzed the conversational strategies used by writers to weaken potential face-threatening speech acts (e.g., requests, orders, accusations). One of these strategies is the addition of lexical softeners, or linguistic hedges, to the message: words and expressions, such as *please* or *maybe*, which mitigate the threatening value of the speech act. Marcoccia & Gauducheau (2007) identified emoticons used with the same purpose as these linguistic hedges and referred to them as "politeness emoji".

Several authors have since studied the use of emoji as linguistic hedges in CMD. Skovholt et al. (2014) investigated a corpus of workplace emails and found emoji used to soften various types of face-threatening acts. In Example 7 below, the "smiling face" emoji is used to weaken a potentially face-threatening order,

whereas in sentence 8 the "winking face" emoji softens a criticism. McCulloch (2019) notes that a message with a full stop at the end tends to be interpreted as more aggressive or negative, and in some kinds of relationships the simple absence of emoji in a message can make a speech act feel more menacing for the receiver (Baron & Ling 2011).

(7) *I'm here, open the door* 😊

(8) *You shouldn't use so many abbreviations in your first paragraph* 😉

While the politeness function is acknowledged by many researchers who studied emoji functions, most typologies tend to include this role within the interpretative function, claiming that politeness emoji also modify the meaning, or the illocutionary force, of the message (Dresner & Herring 2010; Cramer et al. 2016; Spina 2018). However, other linguists consider this use of emoji as closer to the relational function, as relational emoji can function as strengtheners of positive speech acts, just like politeness emoji are softeners of threatening speech acts (Tang & Hew 2019). In order to produce a fine-grained classification of emoji functions, it was decided to view the politeness function as a separate function in this paper.

2.5 The emphatic function

There are two different ways in which emoji can carry an emphatic function. First, emoji can emphasize an emotion or a stance that has already been expressed in the verbal content of the message. By "mirroring" the content (or the illocutionary force) of the message in a redundant way, the emoji can be used to reinforce it (Amaghlobeli 2012; Schneebeli 2017). For instance, in a potentially face-threatening message that already contains verbal hedges, an emoji can be added as another linguistic hedge in order to further soften the message, as in Example 12. Then, in sentence 13, the "star-struck" emoji emphasizes the excitement already expressed in the message by the capital letters and by the elongation in the word "SOOOO". Emoji used for emphasis often carry another linguistic function that they reinforce (Marcoccia & Gauducheau 2007; Skovholt et al. 2014), such as politeness in Example 9 and the expressive function in Example 10.

(9) *Is it okay if I'm 20 minutes late tonight?* 😊

(10) *SOOOO EXCITED FOR THE CONCERT* 🤩

Emphatic emoji can also strengthen the value of other identical emoji in a message (Cramer et al. 2016), like in Example 11 below. An in-depth analysis of data provided by Swiftkey (a virtual keyboard software used on Android and Apple devices) showed that all of the most used combinations of emoji were sequences

of the same emoji, and that more than half of emoji bigrams, trigrams and quadrigrams are repeated emoji (Medlock & McCulloch 2016). Comparing emoji to "beat gestures", McCulloch and Gawne (2018) suggested that the emphatic role played by the repetition of emoji was similar to the emphasis generated by the use of repeated gestures in face-to-face conversation (for example, doing a thumbs-up while moving your hand up and down adds emphasis to the positive value of the thumbs-up). Just like verbal intensifiers (adjectives or adverbs), emphatic emoji only seem to work when linked to another unit which carries the emphasized value, whether this unit is a word or another emoji (Schneebeli 2017).

(11) *My exam was moved to next week* 😭😭😭

2.6 The structural function

Emoji seem to occur in messages at the same position as punctuation marks (i.e., pivotal or final position) most of the time. Early CMD researchers compared their structural function to that of laughter in speech, because laughter is able to "punctuate" face-to-face conversation (Provine et al. 2007). Emoji are often found in final position of messages, similarly to full stops, and can be used to mark the end of a message. To a lesser extent, emoji and emoticons can also be used between two clauses inside a single message (Markman & Oshima 2007). They constitute an efficient way to signal a topic shift inside a conversational turn, as in Example 12 below. Sampietro (2016) observes that messages in which emoji occur almost never contain regular punctuation marks (except for question marks and exclamation marks), suggesting that emoji are efficient substitutes for most punctuation marks in CMC. They can represent a short pause in a sentence or between two sentences, like a comma or a full stop (Panckhurst & Frontini 2020). Emoji can thus be seen as convenient tools for conversational management (Cramer et al. 2016). Like punctuation marks, they are able to work as syntactic markers in a sentence or inside a message, signaling different structural boundaries depending on their position (Amaghlobeli 2012; Spina 2018).

(12) *Yes, I fed the cat* 😊 *what movie do you want to watch tonight?*

2.7 The referential function

The previous functions could almost only be filled by particular categories of emoji. For example, it would be rather unlikely to find an "elephant" emoji or a "bow and arrow" emoji carrying an expressive or a politeness function. Those functions are usually restricted to the "Smileys and People" emoji and to some emoji from the "Symbols" category (i.e., mostly hearts). However, the last func-

tion in this preliminary typology, the referential function, can be assumed by virtually any type of emoji.

(13) *Wanna get some* 🌮 *for lunch?*

(14) 👍 *if you enjoyed the video*

(15) *Always in my* ❤️

(16) *I* ❤️ *you*

Referential emoji, also called "representational" emoji (Beißwenger & Pappert 2019), are used as direct referents of the concept that they represent and thereby replace a lexical word (Amaghlobeli 2012; Cramer et al. 2016). In Example 13, the "taco" emoji stands for the noun "taco" in the sentence, whereas in Example 14 the "thumbs up" emoji replaces the verb "like". Some emoji can replace different parts-of-speech, such as the "red heart" emoji that represents the noun "heart" in Example 15 and the verb "love" in 16. In all these cases, the emoji are used as lexemes on the syntagmatic axis and are part of the propositional content of the message (Schneebeli 2017).

Such emoji can also be used without verbal context, in combination with other referential emoji, in order to create "emoji stories". These narrative sequences are composed of several emoji which, once interpreted together, build a sentence or a "story" (Alshenqeeti 2016; Herring & Dainas 2017). However, emoji stories are a very hard exercise because of the lack of logical and deictic symbols available in the so-called "emoji lexicon" (McCulloch 2019). Example 17, which is an excerpt from Fred Benenson's (2010) attempt to translate Melville's *Moby Dick* into an emoji story, illustrates how hard it is to understand an emoji story without its literal translation. Benenson himself conceded that "there are large portions of the book that don't make any sense" (Goldmark 2014).

(17) 👦🏻💡👍🏻👨🏻🐋🌀💸💸

(I remembered a story of a white man--a whaleman too--who, falling among the cannibals, had been tattooed by them.) (Benenson 2010)

(18) *christmas is almost there!* 🎄⛄🎁

Referential emoji are also found in messages in which the concept represented by the emoji is already present in the verbal content of the message. In Example 18, the three emoji "echo" the referred concept, Christmas, through iconic repetition (Na'aman et al. 2017). Panckhurst & Frontini (2020) refer to those occurrences as "redundant additions", as opposed to emoji used for "necessary addition" and "lexical replacement". Danesi (2016) suggest that these redundant emoji can be used to "avoid potential misunderstandings". Moreover, most digital keyboards on

smartphones and apps include an "emoji prediction" feature that proposes to their user to add, after a word, the corresponding emoji (McCulloch 2019). Besides its referential purpose, this use seems to be able to play expressive, relational, emphatic or structural roles, even though this has not yet been researched.

In concluding, it appears that a single emoji occurrence in a message can carry multiple linguistic functions simultaneously, as stated by Spina (2018). The functions described in this section can be combined in different ways: for example, an interpretative emoji can also have a structural function, and emphatic emoji are always emphasizing another function (mostly expressive). Beißwenger & Pappert (2019) suggest that some functions are "basic functions" while other functions are somewhat secondary. In order to produce the most fine-grained analysis possible, we will try to identify the different functions carried by each emoji inside its message.

3. Methodology

The goal of this paper is to test our preliminary "7-functions" typology, which is intended as a synthesis of prior studies on the subject and as a consensus between those typologies, on a large corpus of emoji used in context. Applying this framework on such a dataset can have several potential outcomes. First, it could prove to be highly successful and demonstrate that the initial typology is totally operational. It could also turn out to be dysfunctional (if the annotation task is too difficult), forcing us to build a completely new framework for the analysis of emoji functions. A third eventuality is that the typology could be only partially successful, with a certain number of flaws, requiring some modifications in order to become operational.

In this section, the corpus used for the study is presented, then the different steps of the annotation and the methods of the analysis are presented.

3.1 Corpus

The corpus on which the preliminary typology of emoji functions was applied contains 1.200 emoji occurrences and the messages (or the "context") in which they were found. They were separated into twelve different subcorpora, as presented in Table 1. Every subcorpus was given an id (from *n1* to *n1200*), and was built in contrast to the other subcorpora on the basis of three different criteria: language of the message in which the emoji was found (English or French), social media on which the message was posted (Facebook or Twitter), and topic of the discussion in which the emoji was used ("Netflix", "Brexit", "Politics", "Everyday

life" and "Everyday life during coronavirus isolation"). Details on each subcorpus can be found in Appendix 1. The emoji in the corpus were collected from contexts following these three criteria equally: 600 emoji were found in French messages and 600 in English, and the same distribution applies to Tweets and Facebook comments (and to the "topic" criterion). This multi-linguistic, multi-platform and multi-topic approach can give to our findings a slight universal aspect (though it has to be confirmed), and it will allow for a comparative analysis of our results in further research.

The decision to create a new corpus, instead of using preexisting corpus such as *88milSMS* (Panckhurst et al. 2016) or *thumbs4science* (Cougnon & Draelants 2018) was motivated by the desire to build as parallel a corpus as possible.

Table 1. Description of the twelve subcorpora

ID	Language	Social media	Topic	Production period	Collection period
CFE	English	Facebook	Everyday life	6/01 to 13/01	10/01 to 13/01
CFF	French	Facebook	Everyday life	9/01 to 16/01	15/01/20 to 16/01/20
NTE	English	Twitter	Netflix	10/01 to 17/01	17/1
NTF	French	Twitter	Netflix	20/01 to 22/01	21/01 to 22/01
NFE	English	Facebook	Netflix	01/02 to 08/02	08/02 to 10/02
NFF	French	Facebook	Netflix	22/01 to 24/01	08/02 to 10/02
PTE	English	Twitter	Brexit	30/01 to 10/02	11/02 to 12/02
PTF	French	Twitter	Brexit	01/01 to 10/02	14/02 to 15/02
PFE	English	Facebook	Politics	20/03 to 24/03	24/03
PFF	French	Facebook	Politics	21/03 to 26/03	25/03 to 26/03
ITE	English	Twitter	Coronavirus isolation	21/03 to 28/03	27/03 to 28/03
ITF	French	Twitter	Coronavirus isolation	28/03 to 29/03	29/03 to 30/03

The French and English-speaking emoji were extracted from Tweets found with various keywords and from comments on public Facebook pages and groups. All emoji and the messages in which they were found were written in 2020 between January 6 and March 30.

For anonymization, all names found in messages included in the corpus were replaced by "John Doe".

3.2 Annotation

Every emoji from each of the twelve subcorpora (along with every message containing each emoji) was placed in an Excel file and was then methodically annotated. Eleven different variables were identified and assigned to each emoji, including the subcorpus in which it belongs, its corresponding code in the Unicode standard (e.g., the "face with tears of joy" emoji, "😂", is *U+1F602*), and the number of times that it was repeated inside the message. Consecutive repetitions of the same emoji were counted as a single occurrence of that emoji (but were almost always annotated as having an emphatic function, induced by the repetition). Annotation guidelines are detailed in Appendix 2.

Then, the position of every emoji in the message was identified and annotated. Five different positions were found to be relevant for classifying emoji, inspired by Uygur-Distexhe's (2012) framework for emoticons positions: initial, between (for "between two clauses"), inside (for "inside a clause"), final and quasi-final. Quasi-final emoji were all followed by "x", "xx" or "xxx", which are typographical renderings of kisses or hugs in CMD (Hilte et al. 2016). These were only found in English Facebook comments, as in Example 19 below.

(19) n51: *You was drunk John Doe doesn't matter if there was another pub* 😜 *xxx*

Then, the most important part of the annotation process was the identification of the emoji's functions, or the application of the preliminary typology of emoji functions outlined in Section 2 to the corpus. Because a single emoji is able to perform more than one function at the same time, up to three different functions were sometimes assigned to one emoji (referred to as *function1*, *function2* and *function3*).

As the identification of an emoji's function(s) can be a very delicate task, it was necessary to verify whether the choices made by the author as the only annotator were sufficiently relevant. Therefore, a second annotator was asked to tag a random sample of 50 corpus examples using the same typology of emoji functions. If the annotator struggled with one of the examples (e.g., due to lack of context), she was given a new random example. Fleiss Kappa was used to compute the inter-rater reliability value, because the annotation consists of more than two categorical variables (Fleiss & Cohen 1973). The test gave the following result: $k = 0.61$, $p < 0.001$. The agreement between the two annotators can thus be interpreted as "substantial" (Gwet 2014).

3.3 Methods

Once the corpus is fully annotated, the frequencies of the functions observed in the corpus are presented and analyzed. The second step is to compute different independence tests in order to examine how the different variables of the corpus influence emoji functions and their frequencies. These variables can be divided into internal (specific to the emoji and the message) and external variables (specific to the user or the context). Internal variables include emoji type (e.g., the type of "😉" is "winking face") and category (e.g., the category of "😉" is "Smileys and People"), as well as emoji position. External variables are the social media, the topic and the language of the message, and the gender of the user (if it was reliably identified during the annotation phase). This paper only focuses on internal variables, as the influence of external variables on emoji functions will be investigated in a future comparative study.

4. Findings

This section presents the results of our corpus analysis. See Section 5 for further discussion.

4.1 Preliminary observations

In the entire corpus, 195 different emoji types (different forms of emoji) were found. Knowing that there were 3,304 emoji types available in March 2020 (month during which the last emoji was integrated in the corpus) in the Unicode Standard, this represents barely 6% of the whole emoji "lexicon".

Table 2 displays the 30 emoji types that were the most frequent in the corpus (excluding emoji repetition) and the number of times each of these was found. It is important to note that 98 of the 195 different emoji types only appear once in the corpus. These single occurrences were referred to as *hapax legomena* (Baayen 2009).

The corpus seems to follow a "power-law" distribution: 66% of the emoji in the corpus belong to one of the 30 emoji in Table 2. The most commonly found emoji, "😂", appears 119 times. It accounts by itself for 10% of the 1200 emoji in the corpus. The second most popular emoji, "🐠", constitutes 5% of the entire data. This suggests that the corpus is in agreement with Zipf's law, according to which the frequency of a given element in the corpus is inversely proportional to its rank in the frequency table (Zipf 1935). This law was already applied by several authors

who found a similar distribution of emoji types in their corpus (Novak et al. 2015; Lu et al. 2016).

Most of the other emoji appearing at the top of the frequency table correspond to those found in the 2019 frequency list revealed by the Unicode Consortium (Unicode 2019): "❤️", "🐠", "😍", "👍" are found really high up on both Unicode's list and our top 30. However, two emoji do not appear among our 30 most frequent emoji, even though they are part of Unicode's top 10: "😘" and "💕". It can be argued that these do not appear as much in our data because all the messages in our corpus were found in public conversations. These two emoji seem to show a high level of intimacy between the scriptor and the recipient and can be expected to occur a lot more frequently in private conversations. Very similar observations can be done when comparing our top ten emoji with those found in the *88milSMS corpus* (Panckhurst et al. 2020).

The high number of "British flag" emoji ("GB") in the corpus is related to the fact that two of the twelve subcorpora revolved around the "Brexit" topic. Besides this emoji and the "red heart" emoji, all the other emoji in Table 2 come from the "Smileys & People" category. In fact, 85% of all emoji in the corpus come from this category, which is not surprising as six of the seven emoji functions are almost exclusively carried by emoji representing faces and attitudes. In addition, 6% of all emoji in the corpus are hearts.

Table 2. The 30 most used emoji types in the corpus, ranked by token frequency

30 most popular emoji in the corpus								
Emoji	**Unicode**	**N**	**Emoji**	**Unicode**	**N**	**Emoji**	**Unicode**	**N**
😂	U+1F602	119	🙂	U+1F642	30	😕	U+2639	15
🐠	U+1F923	64	😊	U+1F60A	27	👏	U+1F44F	13
🤭	U+1F914	52	😅	U+1F605	25	🙏	U+1F64F	13
😍	U+1F60D	51	😘	U+1F622	20	😆	U+1F606	12
😉	U+1F609	48	👌	U+1F44C	19	😒	U+1F60F	10
❤️	U+2764	40	😠	U+1F620	19	👀	U+1F440	10
👍	U+1F44D	38	🤦‍♀️	U+1F937	19	😕	U+1F615	10
😁	U+1F601	34	😱	U+1F631	16	🥰	U+1F970	10
🙄	U+1F644	32	GB	U+1F1EC	16	🧐	U+1F9D0	10
😭	U+1F62D	31	🤦‍♂️	U+1F926	15	🤗	U+1F917	10

Regarding emoji position, it appears that 82.5% of all emoji were found in final position. 11.2% occurred between two clauses, with the other positions being very underrepresented in the corpus. These results are quite similar to the observations that were made on the *88milSMS* corpus by Panckhurst & Frontini (2020). Then, most emoji (82.1%) tend to be used only once in their message, even though 15% are repeated two or three times consecutively in the same message. Two emoji in the corpus were even repeated eight times.

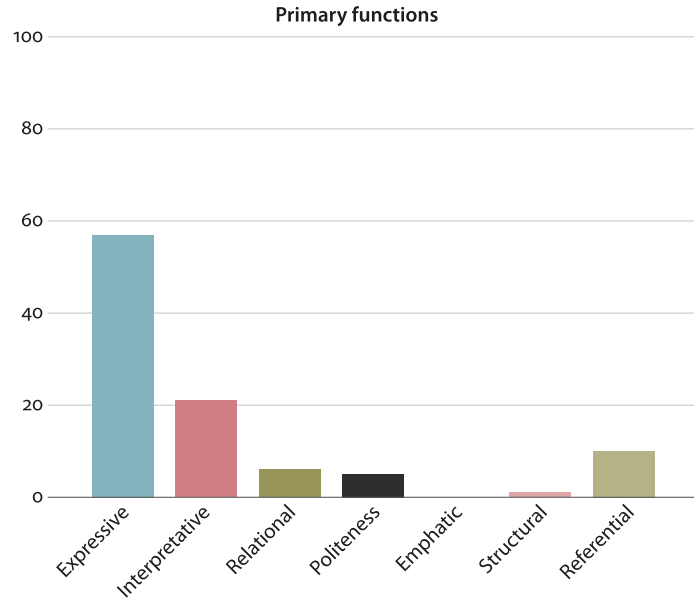
4.2 Distribution of emoji functions

The absolute and relative frequencies of the seven emoji functions are presented in Table 3. Some emoji were identified as having more than one function, which is why the frequencies of secondary and tertiary functions are also listed. Of the 1200 emoji, 67% (804) were identified as having one single function. A secondary function was found for 33% of the emoji, and a tertiary function was assigned to only 2.9% of the data (35 emoji). The primary function of an emoji is by definition its most important function, and thus the one that should be investigated most.

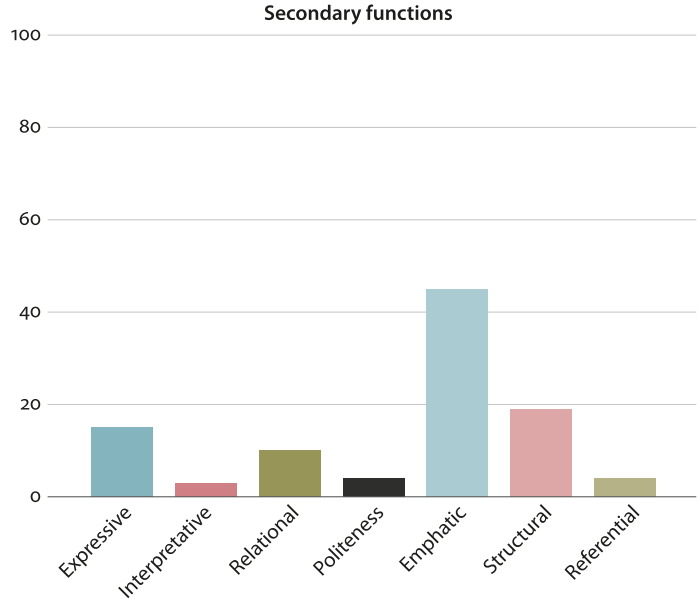
Table 3. Distribution of primary, secondary and tertiary emoji functions in the corpus. Examples of multifunctional emoji are found in Sections 4.2 and 5.1

	Functional level					
	Primary		**Secondary**		**Tertiary**	
Functions	**N**	**%**	**N**	**%**	**N**	**%**
Expressive	687	57.1	57	14.4	2	5.7
Interpretative	253	21.1	13	3.3	0	0.0
Relational	72	5.9	40	10.1	0	0.0
Politeness	55	4.6	15	3.8	0	0.0
Emphatic	2	0.2	179	45.1	18	51.4
Structural	6	0.5	76	19.1	13	37.1
Referential	126	10.5	17	4.3	2	5.7

Figure 1 visualizes the distribution of the primary and secondary functions in the corpus on two pie charts. It is clear that some functions are much more common than others. First, the expressive function has by far the highest frequency of all: more than half of the emoji in the corpus are expressive (687 emoji). When the 59 emoji with a secondary or tertiary expressive function are added, this number rises to 746 (62.1%).



a.



b.

Figure 1. Distribution of primary and secondary functions of emoji in the corpus

The second place is occupied by the interpretative function, which represents more than 20% of the data. Referential emoji represent a little more than 10% of the material. Both of these functions are almost always considered primary. Then, counting primary and secondary functions, a total of 112 relational emoji were found in the corpus. The politeness function was identified for 70 emoji, mostly as a primary function.

The emphatic and structural functions were hardly ever considered to be primary functions. Both are very frequent as secondary and tertiary functions, with 199 emphatic emoji (16.5%) and 95 structural emoji (7.9%) in total. If all three functional levels are taken into account, the emphatic function is the third most common function in the corpus. However, in terms of primary functions only, it is the least frequent.

Next, we investigate how multiple functions combine at different levels for a single emoji. Table 4 shows that, for the 397 emoji with a primary and a secondary function, some combinations are more frequent than others. The most popular combination consists of an expressive primary function with an emphatic secondary function, which occurs 152 times in the corpus (12.6%). This could be explained by suggesting that emotions are the most obvious things that can be emphasized. Example 20 below presents an expressive-emphatic emoji. When expressive emoji are repeated, like the "😂" here, their emotional value is strengthened, and they gain an emphatic function. Consequently, the emphatic function was often assigned to repeated emoji. Interpretative emoji can also have a secondary emphatic function, but this combination is much less frequent (17 cases).

(20) n127: *J'ai 70 ans et Namur à toujours ete désertique le dimanche* 😂😂😂😂
(I'm 70 and Namur has always been desert on Sundays 😂😂😂😂)

The structural function appears to occur at secondary level with all types of primary functions. It is often assigned to emoji that work like punctuation marks (i.e., between two clauses) but that carry another function which is more meaningful. For example, the two "🍺" emoji in message 21 are clearly referential: they refer to the beer(s) that the speaker is going to drink during his "day out". However, the emoji appear between two clauses, exactly where a punctuation mark would have appeared if the emoji had not been there, in order to separate the two clauses. The emoji thus has a secondary structural function.

(21) n1012: "*Day out today* 🍺🍺 *Going hit the beer garden first,(my garden) then go have a disco,(in kitchen)*"

Emoji with three different functions are not very frequent in our corpus, as shown by Table 3 above. As with secondary functions, the emphatic and structural functions are the most common tertiary functions carried by emoji.

Table 4. Primary and secondary functions of the emoji in the corpus. Columns are the primary functions of emoji, rows are the secondary functions

	Primary						
Secondary	**Expressive**	**Interpretative**	**Relational**	**Politeness**	**Emphatic**	**Structural**	**Referential**
Expressive	///	20	20	4	1	2	10
Interpretative	8	///	1	4	0	0	0
Relational	34	3	///	2	0	0	1
Politeness	10	4	0	///	0	0	1
Emphatic	152	16	0	0	///	0	11
Structural	41	17	3	4	0	///	11
Referential	12	3	0	1	0	1	///
None	430	190	48	40	1	3	82

$X^2 = 262.43^{***}$, $p < 2.2e\text{-}16$

4.3 The influence of internal factors

4.3.1 *Emoji types*

As seen in Table 5, some emoji types appear to occur very often with the same function. Of the ten most frequent emoji types found in the corpus, four emoji are almost always primarily expressive: "😭", "😍", "🤭" and "👍". The most frequent emoji, "😂", is also often used with an expressive function, but even more with an interpretative function. The second most popular emoji, "💸", can play these two roles as well. For example, in message 22 below, the three emoji are primarily expressive, because they are used to represent the speaker's feelings about the movie. Then, in Example 23, the emoji are used with a primary interpretative function (they show irony). Those two most frequent emoji are very similar in appearance, but the main difference between them seems to be functional: "💸" is more likely to be interpretative (62.5%) than "😂" (52%).

(22) n298: *Mindhorn 2017 #Netflix What a funny movie… loved it*❤️💸💸

(23) n179: *s'est sur que le train avec les sacs de courses s'est pratique.. S'est bien connu* 😂😂😂

(sure, taking the train with shopping bags is convenient… It's well known 😂😂😂)

Another interesting emoji is the "winking face" emoji ("😉"). It is very frequently used for displaying irony (interpretative function). However, it also frequently carries two other functions, the relational and the politeness functions. The emoji

in Examples 24 and 25 illustrate this: the first "😉" is used by the speaker for maintaining a positive relationship with the addressee (relational), and the second one acts as a linguistic hedge, mitigating the potentially face-threatening value of the message (politeness).

(24) n948: *Bonsoir bonne lecture et bonne soirée* 😉
(Good evening, have a pleasant read and a good night 😉)

(25) n195: *c'est sans doute comme ça que vous faites mais ne vous en déplaise, ce n'est est pas comparable. Je respecte votre opinion mais je ne la partage absolument pas* 😉
(it's probably how you do it but whether you like it or not, it is not comparable. I respect your opinion but I do not share it at all 😉)

Table 5. The distribution of primary functions for the 10 most popular emoji types (by token frequency)

	10 most frequent emoji types									
P. functions	😂	🎉	🤭	😍	😉	❤️	👍	😁	🙄	😭
Expressive	55	24	49	50	1	28	35	15	20	31
Interpretative	62	40	3	0	22	1	0	14	11	0
Relational	0	0	0	1	15	8	3	3	0	0
Politeness	2	0	0	0	10	0	0	1	1	0
Emphatic	0	0	0	0	0	0	0	1	0	0
Structural	0	0	0	0	0	0	0	0	0	0
Referential	0	0	0	0	0	3	0	0	0	0

$X^2 = 3990.6^{***}$, $p < 2.2^{e\text{-}16}$

Even though many emoji types are intrinsically linked to a single specific function, the results show that some kinds of emoji can be used for different functions. The most popular emoji types logically appear to be the ones which can fulfil a wider variety of linguistic functions in conversation.

4.3.2 *Emoji categories*

Regarding emoji categories, our findings show that, as expected, the first six functions of the preliminary typology can almost only be carried by "Smileys & People" emoji. Furthermore, referential emoji are mostly from all the other categories (e.g., animals, food, objects). However, an unexpected finding concerns the "Objects" emoji, which sometimes appear to have an expressive function. After investigation, it appears that this is related to the presence of a particular emoji

from the "Objects" category in the corpus: the "party popper" emoji ("🎉"). As shown by Example 26 below, in which the speaker is referring to the release date of a popular movie on Netflix, this emoji is often used to convey excitement and, in other contexts, to express congratulations. This "metonymical" use of the emoji relates to that of the "fire" emoji ("🔥"), which is also frequently used to express excitement and not to refer to proper fire.

(26) *12th feb will be a happy happy day* 🎉

4.3.3 *Emoji positions*

Table 6. Influence of emoji position on primary function

	Emoji position					
Primary function	**Final**	**Between**	**Inside**	**Initial**	**Quasi-final**	**Between & final**
Expressive	585	67	5	17	7	4
Interpretative	215	32	1	0	3	2
Relational	64	5	1	5	1	0
Politeness	49	6	0	0	0	0
Emphatic	2	0	0	0	0	0
Structural	0	3	0	3	0	0
Referential	77	22	21	0	0	1

$X^2 = 1443.7^{***}$, $p < 2.2^{e\text{-}16}$

The position of an emoji appears to have a significant effect on its function. One clear finding that can be inferred from Table 6 is that all primary functions (except for structural and emphatic) are mainly found with emoji in final position. Interestingly, many of the emoji in initial and quasi-final position are expressive, but never referential. On the other hand, emoji found inside a clause are frequently identified as having a primary referential function. This makes a lot of sense, as many of the emoji placed inside the syntactic structure of a clause are there because they are used to replace a noun or a verb.

5. Discussion

5.1 Review of the initial typology

The annotation and analysis of the corpus provided interesting findings on the different linguistic functions of emoji in conversation. In this section, after investi-

gating every function in the light of the results obtained, we will revise our initial framework in order to build an updated version of our initial typology.

It is not surprising that the most natural function of emoji is also the most frequent in our corpus, with a total of 746 emoji (62.1% of the data) identified as expressive at some level (yet mainly at primary level). These results support those of previous studies on the frequencies of emoji functions (Cramer et al. 2016; Herring & Dainas 2017). The expressive function is almost only assigned to facial emoji, gesture emoji and hearts. However, other particular types of emoji have been analyzed as expressive as well (e.g., "🔥" and "🎉"). Overall, the expressive function is unequivocally the most important linguistic function of emoji in context. It can be used in various situations to represent many different emotions (of both the speaker and of other individuals mentioned by the speaker). Therefore, it should keep its predominant place inside our revised typology of emoji functions.

As in the results of the study carried out by Cramer et al. (2016), the second most frequent emoji function in our corpus is the interpretative one, which has been identified for 266 emoji (22.2% of the data). This function is almost always considered to be a primary function, sometimes associated with other secondary functions (mostly expressive and structural). The only emoji types capable of altering the meaning of the message or signaling irony appear to be facial emoji, except for a few gesture emoji, and the "🙈" emoji, which was discussed in Section 3.1.2. The "😂" and "🤣" emoji are the most common emoji types in the entire corpus, but they are also the most frequent interpretative emoji. Some emoji in the corpus were considered as interpretative even though they do not signal irony or do not alter the meaning of the message, which were the only two features assigned to interpretative emoji in the initial typology. In Example 27, the "smirking face" emoji ("😏") is placed at the end of the message. However, the content of the message is not ironic (as the writer does not state the opposite of what he actually means; the emoji is used here in order to alert the reader to an implicature in the message. According to Grice's theory of conversation, an implicature is something that is implied by the speaker in an utterance, but that is not literally expressed in that utterance (Grice 1975). In this example, the speaker's implicature is: "what Miller says does not make sense". It could thus be argued that the role of interpretative emoji is to signal that the message contains an implicature, which includes ironical utterances. This should be considered when reviewing our typology, in which the interpretative function definitely has to be considered as one of the most important linguistic functions of emoji.

(27) *n738: "Donc pour #Miller les Anglais ont quitté l'Europe de #Macron …en 2016 vote du #Brexit et #Macron a été élu en 2017 ..* 😎😒*" (talking about Gerard Miller, French TV commentator)*
("So for #Miller the British have left #Macron's Europe… in 2016, #Brexit was voted, and #Macron was elected in 2017.. 😎😒")

Relational emoji were not very frequent in the corpus. This is probably linked to the fact that our corpus only contains public or semi-public Tweets and Facebook comments, and that phatic interactions (e.g. greetings or small talk) are more likely to occur in private conversations. Yet, 72 emoji were identified as primarily relational, and 40 as relational at a secondary level. Relational emoji are always emoji types with an essential "positive" meaning, like "🥰", "✌" or simply "🙂", because they are used by the speaker to maintain a positive relationship with the reader. An interesting observation is that relational emoji often have an expressive function as well, at primary or secondary level. In total, 85% of secondary relational emoji are primarily expressive, and the secondary expressive function was assigned 20 times to primary relational emoji. Hence, it may be necessary to reconsider the importance of the relational function in our typology.

Only 70 emoji in the whole corpus were identified as politeness emoji. For the majority of these (78.5%), politeness was the primary function. They were mostly found in final position, and almost only assigned to facial emoji: "😉", "😊" and "🙂" were the most frequent emoji types found with the politeness function. It can be observed that many of the emoji types associated with politeness have a "weakly positive value" (such as the "slightly smiling face" emoji, "🙂"). This is likely due to the fact that politeness emoji are used to mitigate the potential face-threatening value of the content of the message. They constitute small "positive" softeners in generally "negative" messages. They are mostly used as linguistic hedges in situations where the face of the receiver is potentially threatened. These include orders, disapprovals or criticisms, as in Example 28.

(28) *n935: "Allez, essayez de réfléchir sur ladite "Morale", sur les conventions-sociales et les (soi-disantes) bonnes moeurs.* 😉*"*
("Come on, try to think about said "morality", about social conventions and the (so-called) good behaviors. 😉")

Emoji were only assigned the emphatic function as a secondary or tertiary function, and it was almost only assigned to primarily expressive emoji. The very high frequency of expressive-emphatic emoji is due to the fact that emotions are the most natural thing that can be emphasized in discourse. The emphatic function can be produced through emoji repetition, through emoji combination (reinforcing the value of a different emoji), or by mimicking an emotion that is already expressed in the textual content of the message.

It appears that the particularity of the structural function is that it can be combined with any other emoji function, and that it can be applied to virtually every emoji type. In fact, it seems to be very much dependent on emoji position: 75.7% of all structural emoji were found between two clauses. This is due to the fact that most structural emoji are used to separate two clauses inside a message, just like punctuation marks (mainly commas and full stops). The structural function is almost always secondary, as the emoji separating clauses usually have another primary function. Like the emphatic function, it is clear that the structural function should not be considered as a primary emoji function in our revised typology. It works as a supplementary or subordinate function that all emoji types can carry.

Finally, the referential function is the third most frequent function in our corpus, following the expressive and the interpretative functions, which tends to confirm the results found by Herring & Dainas (2017) regarding frequencies of emoji functions. This function appears to be the default function of all emoji types that do not belong in the "Smileys & People" category. This is quite natural, as all the other emoji types are used to represent items, animals or activities that are technically unable to convey anything other than the iconic concept they refer to. For example, the "🍎" emoji is typically only used to refer to the concept "apple". Still, there are some exceptions (e.g "🎉", "🔥" or "🙈") to this rule. Referential emoji are mostly found in final position, but they can also occur inside a clause, when they are used to replace a word. Indeed, they can be used to replace nouns or noun phrases, but also verbs, adjectives or even adverbial phrases, as in Example 29 below. Other referential emoji can also refer to words or concepts that are already present in the verbal content of the message. They can also evoke a concept that is not explicitly mentioned in the message but that adds some contextual information to it. For instance, the "💃🏻" emoji in Example 30 is used by the speaker to imply that she was dancing on Wednesday nights. These particular referential emoji could be considered in some sense as carrying an interpretative function, because they help the reader understand the implied meaning of the message. Overall, the referential function appears to be quite frequent, and should clearly be considered as one of the main emoji functions, alongside the expressive and interpretative functions. Moreover, referential emoji can also carry other functions simultaneously.

(29) *n500: "it's so overrated rip, the kissing booth is 💯 better"* (a hundred times better)

(30) *n71*: "Wednesday nights 👌🏻💃🏻"

Then, it can also be argued that emoji, in particular referential ones, add some sort of aesthetic value to the message in which they are produced. Emoji can have

a decorative purpose, making the message more colorful and more visually attractive to potential readers. In addition to this, producing and reading emoji can be a source of enjoyment for some users, which also contributes to this "aesthetic" or "ludic" function. However, it would not be reasonable to say that all emoji are produced with an aesthetic purpose. Yet in some messages, it is unclear what the actual purpose of the emoji is, other than the addition of "decoration" to the message. Even though this aesthetic function was not considered in our initial typology, it appears from our analysis that it should be included in the revised classification of emoji functions presented below.

5.2 A revised typology of emoji functions

Table 7. The revised typology of emoji functions, with examples

Primary functions	Expressive	Conveying emotion, whether the speaker's own emotional state or someone else's. *n20*: ""my aunty and uncle sadly passed 😔"
	Interpretative	Signaling the presence of an implicature in the message and guiding the interpretation of the implicature (including ironical and sarcastic utterances). *n1082*: "Because we'd all love a visit to A&E right now 😕"
	Referential	Referring to a concept inside or outside the message, by replacement, evocation or repetition. *n809*: "I also have a unicorn 🦄."
Secondary functions	Relational	Maintaining a positive relationship with the reader(s) of the message. Combined with expressive or referential emoji. *n5*: "Thank you very much 🙂"
	Politeness	Softening the potential face-threatening value of a message. Combined with expressive and interpretative emoji (or referential emoji for weakened insults). *n819*: "I did. Read my comment 😉"
	Emphatic	Strengthening the emotional value of a message or an emoji. Combined with expressive emoji (and politeness emoji). *n219*: "Just finished watching #TheSkyIsPink 😭😭😭😭😭😭"
	Structural	Separating, framing and punctuating clauses. Combined with expressive, interpretative and referential emoji. *n82*: "I started there 63 or 64 😂 loved it"
	Aesthetic	Adding a decorative value to the emoji, making the message more attractive, more colorful or playful. Mainly combined with referential emoji. *n1037*: "Vegan pizza! 🍕🌱🍕🌱🍕🌱"

The first and most important adjustment that was made to the typology is to separate the functions into two categories: primary and secondary functions. In the initial framework, all seven functions were considered to be distinct and at the same level. However, during the annotation phase, it emerged that some emoji could carry multiple functions at the same time. Some functions were almost always considered to be primary (i.e., the interpretative and referential functions), while others were always seen as secondary (i.e., the emphatic and structural functions). Therefore, in this revised typology, it was decided to identify essentially primary and secondary functions. The five secondary functions in the typology can only be assigned to emoji which already have one of the three primary functions (expressive, interpretative or referential). For example, an emoji can only be classified as relational if it is primarily expressive. However, this does not prevent an emoji from having two primary functions, or two secondary functions, simultaneously.

Then, as suggested in our hypothesis, a new important function of emoji has been identified and added to the new typology. The aesthetic function appears to be a major function fulfilled by many emoji, which was wrongly excluded from our initial typology. Besides that, it was decided to keep all the other functions in the typology, as they all appeared quite frequently throughout the corpus. A total of eight different functions – three primary and five secondary functions – compose this new typology presented in Table 7.

6. Conclusion

Our revised typology of emoji functions arose from the analysis of a corpus composed of 1200 emoji in context, as well as from previous studies and other authors' typology proposals. Seven functions were found to be relevant, and an eight function was identified during the analysis: the aesthetic function. It was also decided to update our initial typology as a two-leveled typology, with three primary functions (expressive, interpretative and referential) and five secondary functions (relational, politeness, emphatic, structural and aesthetic). Some functions occurred more often than others throughout the corpus. The most frequent emoji function is the expressive function (746 emoji) and the least frequent is the politeness function (70 emoji).

The emoji used for the corpus analysis were found in French and English Tweets and Facebook comments on a variety of topics. The typology has been proved to be applicable for the analysis of emoji across two platforms and two languages and can thus be used to analyze emoji produced on the same social networking sites and in the same languages. Yet, it can be suggested that this

framework could also be operational in other similar contexts of CMC and in some other languages, though this hypothesis would have to be verified through further research.

Even though our analysis led to interesting results, there are some inherent limitations which have to be taken into consideration. First, it should be noted that the functions of each of the 1200 emoji in the corpus were identified and analyzed by someone who did not produce these emoji. Obviously, it is not possible to know exactly what was going on in the mind of every user when they produced the emoji, and if the precise functions intended by each user for each emoji were exactly the same as those identified by the annotator. Analyzing every emoji one by one was very time-consuming, because it had to be done very carefully in order to avoid misinterpretations. Therefore, when it turned out too difficult to identify an emoji's linguistic function(s), that particular emoji was excluded from the corpus (most of the times due to lack of context). In addition, a second annotator analyzed a sample of the corpus in order to compute an inter-rater reliability agreement, which showed that the annotation of the 1200 emoji in the corpus could be considered trustworthy. Yet, although the risks of misinterpretation were minimized, it is impossible to know exactly what function was intended by a writer (except by asking them directly), so some emoji in the corpus may have been misinterpreted.

Then, it has to be pointed out that the devices used for writing every message in the corpus, and thus every emoji, are unknown. It is impossible to determine whether a given emoji was produced, for example, from a Mac computer or from an Android Smartphone. The exact appearances of emoji can vary from one keyboard to the other, which may lead to some misunderstandings between users (Miller et al. 2016). Moreover, some smartphones have "smart keyboards", which can predict and suggest emoji based on the words written by their user. Therefore, it is conceivable that some of the emoji in the corpus may have been added to the message independently of the speakers' will. However, this is not an important limitation, as emoji can easily be deleted before publication, and because Facebook comments can be edited after publication. Leaving the emoji in the message is thus still part of the user's decision.

Finally, a more serious limitation is the influence of time on emoji production. More than any other kind of human language, computer-mediated discourse evolves rapidly (Herring 2019). Emoji have been around for more than a decade, but that does not mean that they will still be used as much ten years from now. Even if they stick around, their linguistic functions and the way people use them might be completely different in some time. Furthermore, new emoji enter the Unicode Standard every year, and some emoji become more popular due to short or long-time trends. For example, the "" emoji gained a boost of popularity due

to the coronavirus crisis. For these reasons, it is possible that our final typology of emoji functions and our findings will become obsolete in the future. The emoji analyzed in this paper were produced between January 6, 2020, and March 30, 2020, over a period of three months. The results of our analysis can thus be considered as valid for the first quarter of the year 2020, and it could be hypothesized that they would still be relevant after a few months or even years. However, it would be interesting to produce the same investigation after some time, or to compare our results with older CMC corpora, in order to verify how much the linguistic functions of emoji are influenced by the time period during which they are produced.

References

Alshenqeeti, H. 2016. Are Emojis Creating a New or Old Visual Language for New Generations? A Socio-semiotic Study. *Advances in language and Literary Studies*, *7*: 6, 56–69.

Amaghlobeli, N. 2012. Linguistic features of typographic emoticons in SMS discourse. *Theory and Practice in Language Studies*, *2*: 2, 348–354. https://doi.org/10.4304/tpls.2.2.348-354

Aull, B. 2019. A study of phatic emoji use in WhatsApp communication. *Internet Pragmatics*, 1–34. John Benjamins Publishing Company. https://doi.org/10.1075/ip.00029.aul

Austin, J.L. 1962. *How to do things with words.* (*2nd Ed.*) Edited by J.O. Urmson & M. Sbisà. Cambridge, MA: Harvard University Press.

Baayen, R.H. 2009. Corpus linguistics in morphology: morphological productivity. *Corpus linguistics. An international handbook*, 900–919.

Baron, N. & Ling, R. 2011. Necessary Smileys & Useless Periods: redefining punctuation in electronically-mediated communication. *Visible Language*, *45*, 45–67.

Beißwenger, M. & Pappert, S. 2019. How to be polite with emojis: a pragmatic analysis of face work strategies in an online learning environment. *European Journal of Applied Linguistics*, *7*: 2, 225–253. https://doi.org/10.1515/eujal-2019-0003

Benenson, F. 2010. *Emoji Dick*, Lulu.

Brown, P. & Levinson, S.C. 1987. *Politeness. Some universals in language usage*. Cambridge: University Press. https://doi.org/10.1017/CBO9780511813085

Cougnon, L.A. & Draelants, H. 2018. Language Ideologies and Writing Systems in CMC: a Sociolinguistic Approach. *Language and the new (instant) media, Cahiers du Cental*, *9*, 88–99.

Cramer, H., de Juan, P. & Tetreault, J. 2016. Sender-Intended Functions of Emojis in US Messaging. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 504–509. https://doi.org/10.1145/2935334.2935370

Danesi, M. 2016. *The semiotics of emoji: The rise of visual language in the age of the internet.* Bloomsbury Publishing.

Dresner, E. & Herring, S. 2010. Functions of the Non-Verbal in CMC: Emoticons and Illocutionnary Force. *Communication Theory*, *20*, 249–268. https://doi.org/10.1111/j.1468-2885.2010.01362.x

Dürscheid, C. & Siever, C.M. 2017. Jenseits des Alphabets–Kommunikation mit Emojis. *Zeitschrift für germanistische Linguistik*, *45*: 2, 256–285. https://doi.org/10.1515/zgl-2017-0013

Escouflaire, L. 2020. Building a typology of the linguistic functions of emoji: a cross-linguistic and cross-platform corpus analysis of emoji in conversation (Master's thesis). Supervised by Hilde Hasselgård (UiO) and Cédrick Fairon (UCLouvain).

Fleiss, J.L. & Cohen, J. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*, 613–619. https://doi.org/10.1177/001316447303300309

Fullwood, C., Quinn, S., Chen-Wilson, J., Chadwick, D. & Reynolds, K. 2015. Put on a smiley face: Textspeak and personality perceptions. *Cyberpsychology, Behavior, and Social Networking*, *18*: 3, 147–151. https://doi.org/10.1089/cyber.2014.0463

González-Ibáñez, R., Muresan, S. & Wacholder, N. 2011. Identifying sarcasm in Twitter: a closer look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers (ACL-2011)*.

Grice, P. 1975. Logic and conversation. In P. Cole & J. Morgan (Eds.). *Syntax and semantics. 3: Speech acts*, 41–58. New York: Academic Press.

Gumperz, J. 1982. *Discourse strategies*. Cambridge, England: CUP. https://doi.org/10.1017/CBO9780511611834

Gwet, K.L. 2014. *Handbook of Inter-Rater Reliability (4th Edition). Chapter 6*. Gaithersburg: Advanced Analytics, LLC.

Herring, S.C. (Ed.) 1996. *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*, 81–106. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.39.08her

Herring, S.C. 2001. Computer-mediated discourse. In D. Tannen, D. Schiffrin, & H. Hamilton (Eds.), *Handbook of Discourse Analysis*, 612–634. Oxford: Blackwell.

Herring, S.C., & Dainas, A. 2017. "Nice picture comment!" Graphicons in Facebook comment threads, *Proceedings of the 50th Hawaii international conference on system sciences*, 2185–2194. Waikoloa, Hawaii. https://doi.org/10.24251/HICSS.2017.264

Hilte, L., Vandekerckhove, R., & Daelemans, W. 2016. Expressiveness in Flemish online teenage talk: A corpus-based analysis of social and medium-related linguistic variation. *Proceedings of the 4th Conference on CMC and Social Media Corpora for the Humanities*, 30–33.

Janssen, J.H., IJsselsteijn, W.A. & Westerink, J.H.D.M. 2014. How affective technologies can influence intimate interactions and improve social connectedness. *International Journal of Human-Computer Studies*, *721*, 33–43. https://doi.org/10.1016/j.ijhcs.2013.09.007

Kelly, R. & Watts, L. 2015. Characterising the Inventive Appropriation of Emoji as Relationally Meaningful in Mediated Close Personal Relationships. *Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design*.

Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., & Mei, Q. 2016. Learning from the ubiquitous language: an empirical analysis of emoji usage of smartphone users. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 770–780. https://doi.org/10.1145/2971648.2971724

Marcoccia, M., & Gauducheau, N. 2007. L'analyse du rôle des smileys en production et en réception: un retour sur la question de l'oralité des écrits numériques. *Glottopol. Revue de sociolinguistique en ligne*, *10*, 38–55.

Markman, K. M. & Oshima, S. 2007. Pragmatic play? Some possible functions of English emoticons and Japanese kaomoji in computer-mediated discourse. In *Association of Internet Researchers Annual Conference*, *8*.

McCulloch, G. 2019. *Because Internet. Understanding the new rules of language*. New York City: Riverhead Books.

McCulloch, G. & Gawne, L. 2018. Emoji grammar as beat gestures. *Proceedings of the First International Workshop on Emoji Understanding and Applications in Social Media*, 1–4.

Medlock, B. & McCulloch, G. 2016. *The Linguistic Secrets Found in Billions of Emoji*. Presented at SXSW, March 11–20, 2016. Retrieved 19 October 2019 from https://fr.slideshare.net/SwiftKey/the-linguistic-secrets-found-in-billions-of-emoji-sxsw-2016-presentation-59956212

Na'aman, N., Provenza, H. & Montoya, O. 2017. MojiSem: Varying linguistic purposes of emoji in (Twitter) context. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics – Student Research Workshop*, 136–141.

Novak, P. K., Smailović, J., Sluban, B. & Mozetič, I. 2015. Sentiment of Emojis. *PloS One*, *10*(12).

Panckhurst, R., Détrie, C., Lopez, C., Moïse, C., Roche, M. & Verine, B. 2016. 88milSMS. A corpus of authentic text messages in French. *Banque de corpus CoMeRe*. Ortolang: Nancy.

Panckhurst, R. & Frontini, F. 2020. 4 Evolving interactional practices of emoji in text messages. In *Visualizing Digital Discourse*, 81–104. De Gruyter Mouton. https://doi.org/10.1515/9781501510113-005

Panckhurst, R., Lopez, C. & Roche, M. 2020. A French text-message corpus: 88milSMS. Synthesis and usage. *Corpus*, *20*. https://doi.org/10.4000/corpus.4852

Provine, R. R., Spencer, R., & Mandell, D. 2007. Emotional expression online: Emoticons punctuate website text messages. *Journal of Language and Social Psychology*, *26*: 3, 299–307. https://doi.org/10.1177/0261927X06303481

Sampietro, A. 2016. Exploring the punctuating effect of emoji in Spanish Whatsapp chats. *Lenguas Modernas*, *47*, 91–113.

Schneebeli, C. 2017. The interplay of emojis, emoticons and verbal modalities in CMC: a case study of YouTube comments. *VINM 2017: visualizing (in) the new media*.

Searle, J. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Skovholt, K., Grønning, A. & Kankaanranta, A. 2014. The Communicative Functions of Emoticons in Workplace E-Mails: :-). *Journal of Computer-Mediated Communication*, *19*, 780–797. https://doi.org/10.1111/jcc4.12063

Spina, S. 2018. Role of Emoticons as Structural Markers in Twitter Interactions. *Discourse Processes*, *56*: 4, 345–362. https://doi.org/10.1080/0163853X.2018.1510654

Steinmetz, K. 2015 (November 16). Oxford's 2015 Word of the Year Is This Emoji. *Time*. Retrieved 15 January 2021 from https://time.com/4114886/oxford-word-of-the-year-2015-emoji/

Tang, Y. & Hew, K. F. 2019. Emoticon, Emoji, and Sticker Use in Computer-Mediated Communication: A Review of Theories and Research Findings. *International Journal of Communication*, *13*, 2457–2483.

Thurlow, C., Dürscheid, C. & Diémoz, F. 2020. Visualizing Digital Discourse. Interactional, Institutional and Ideological Perspectives. *Language and Social Life [LSL]*, *21*, De Gruyter Mouton

Troiano, G. & Nante, N. 2018. Emoji: What does the scientific literature say about them? A new way to communicate in the 21st century. *Journal of Human Behavior in the Social Environment*, *28*: 4, 528–533. https://doi.org/10.1080/10911359.2018.1437103

Unicode. 2019 (October 18). *Emoji frequency. The most frequent emoji*. Retrieved 16 September 2021 from https://home.unicode.org/emoji/emoji-frequency/

Uygur-Distexhe, D. 2012. Lol, mdr and ptdr. An inclusive and gradual approach to discourse markers. *Lingvisticæ Investigationes*, *35*: 2, 389–413. https://doi.org/10.1075/li.35.2.13uyg

Weissman, B. & Tanner, D. 2018. A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension. *PLoS ONE 13*(8). https://doi.org/10.1371/journal.pone.0201727

Zipf, G. 1935. *The psychobiology of language*. Boston: Houghton Miffin.

Appendix 1. Content of the 12 subcorpora

1. CFE: Facebook comments from a group containing around 10,000 people, with 40 posts per day in average. Most users are inhabitants of a British city north of Manchester, with 75,000 inhabitants. On this group, people react to old pictures of the city, comment on events happening around town and discuss the local football team's latest matches.
2. CFF: Facebook comments and posts from a group containing around 10,000 people, with 30 posts per day in average. Most users are inhabitants of a Belgian city with 110,000 inhabitants in the French-speaking part of the country. Users react to pictures of the surrounding nature, debate on the width of bicycle paths in the center and argue on which of the city's cafés has the best terrace.
3. NTE: English-language Tweets containing the hashtag *#Netflix*. Users comment on movies and shows that have been announced or that they have watched (e.g. the "Studio Ghibli" movies or the Netflix-exclusive shows *Sex Education* and *The Witcher*).
4. NTF: French-language Tweets containing the hashtag *#Netflix*. Users comment mostly on the same movies and shows as English-speaking users.
5. NFE: Facebook comments on a post from the official "Netflix UK" Facebook page, announcing the new movies and shows coming out in February on the streaming platform. Users react to the announcement, tag their friends to share their excitement and complain about the shows that were not mentioned.
6. NFF: Facebook comments on a post from the French official "Netflix" Facebook page. The post announced the new movies and shows coming out in February on the streaming platform. French-speaking users' reactions were similar to the reactions of the British users.
7. PTE: English-language Tweets containing the hashtag *#Brexit*. Some users praise the Brexit and Boris Johnson, others criticize them. Some speakers of the subcorpus are British politicians.
8. PTF: French-language Tweets containing the hashtag *#Brexit*. Some users criticize the Brexit and Boris Johnson, others support them (often asking for a "Frexit"). Some speakers of the subcorpus are French politicians.
9. PFE: Facebook comments and posts from a group about UK politics containing 6,500 users, and comments from the public page "UK politics", followed by 325,000 people. Users argue over political events and decisions in the UK, mostly over the Brexit, Boris Johnson and the coronavirus crisis.

10. PFF: Facebook comments and posts from the public groups "France: Débats sur la politique" (15,000 members) and "Politique" (14,500 members), and the newspaper's Facebook page "Le Monde Politique" (followed by 265,000 users). People debate on French politics, mostly on political decisions regarding the coronavirus crisis.
11. ITE: English-language Tweets containing the hashtag *#IsolationLife*. Users talk about their life during the forced isolation period provoked by the coronavirus pandemic. For example, they share ideas of things to do to keep busy and react to other people's Tweets.
12. ITF: French-language Tweets containing the hashtags *#ConfinementJour12* and *#ConfinementJour13*. Users talk about their daily lives during the forced isolation period provoked by the coronavirus pandemic, as in the ITE subcorpus.

Appendix 2. Annotation variables

A. *number*
→ The number assigned to the analyzed emoji, from *n1* to *n1200*. These numbers were assigned in a temporal order: the first emoji that was collected is *n1*, the second one *n2*, etc. They represent the "ID number" of the emoji and will allow us to refer to specific emoji and messages in examples further in this paper.
B. *message*
→ The "raw" message in which the emoji was found. This message can be a Tweet, a Facebook post or a Facebook comment.
C. *message_cleared*
→ The same message, but only keeping the parts that are necessary to the analysis of the emoji. The pieces that can be removed from the raw message include hashtags (e.g. *#brexit*), mentions and tags (e.g. *@realDonaldTrump*), as well as words and sentences that are irrelevant to the analysis of the emoji.
D. *platform*
→ The platform on which the message containing the emoji was posted: annotated as *facebook* or *twitter*.
E. *language*
→ The language in which the message containing the emoji was written: *fr* (French) or *en* (English).
F. *subcorpus*
→ The message thread or Facebook group in which the message containing the emoji was found. This corresponds to the subcorpus in which the message is placed (as detailed in Table 2 in Section 2.2.2.).
G. *type_emoji*
→ The type of emoji used in the message, from the 3,304 different emoji types available in the Unicode Standard. It refers to the general appearance of the emoji, independently of the potential visual differences that may exist from one platform to another. For example, the "pizza" emoji, "🍕", contains four slices of pepperoni on Apple devices, while its Twitter appearance has six slices, and the Facebook model only has three. The differences are often minimal, but they exist, and can cause potential confusion when received on a different platform than the one they were sent from (Miller et al., 2016). The appearances used for every emoji in this paper are the Apple models: they are the ones primarily used

on Emojipedia, and they constitute a common ground between Facebook's and Twitter's emoji appearances, allowing for easier comparison of data.

H. *unicode*
→ The code corresponding to the type of emoji, according to the Unicode standard (e.g. *U+1F602* for the "face with tears of joy" emoji, "😂").

I. *category*
→ The category in which the emoji belongs, from the eight different emoji categories listed on most smartphones and on Emojipedia: *people* (for "Smileys & People"), *animals* (for "Animals & Nature"), *food* (for "Food & Drink"), *activity, travel* (for "Travel & Places"), *objects, symbols* and *flags*.

J. *repetition*
→ The number of times that the emoji is repeated (with the same function and at the same position) inside the message. The numbers vary from 1 to 8, with the default amount being 1 (when the emoji is not repeated and thus only appears once inside the message).

K. *same_message* (boolean)
→ Whether the emoji appears in the same message as the previous emoji or not. For instance, in Example 1 below, the emoji "😳" (*n672*) is annotated as "FALSE", but the emoji "👿" (*n673*) is tagged as "TRUE" because it was found in the same message as the previous emoji. This allows for the investigation of emoji combinations as well as individual uses of emoji.

Address for correspondence

Louis Escouflaire
Rue Haute Pensée 11A
7830 Silly
Belgium
louis.escouflaire@student.uclouvain.be

Publication history

Date received: 27 March 2021
Date accepted: 24 October 2021

Author Queries

- Please provide a complete reference for the citation '(Miller et al., 2016), (Herring 2019), (Herring & Dainas 2017), (Goldmark 2014)' in this article.
- Please provide a citation for the reference id "CIT0015 (Escouflaire, L. 2020), CIT0020 (Gumperz, J. 1982), CIT0048 (Thurlow, C., Dürscheid, C. & Diémoz, F. 2020)" since citation is missing in the article.