UiO : Universitetet i Oslo

UCLouvain

Building a typology of the linguistic functions of emoji

A cross-linguistic and cross-platform corpus analysis of emoji in conversation

MA Dissertation
Louis Escouflaire

Supervisors

Hilde Hasselgård
Universitetet i Oslo

Cédrick Fairon
Université catholique de Louvain

Master in Linguistics
Linguistics and its Applications for a Multilingual Society
Double Degree Programme

Academic Year 2019 – 2020

UNIVERSITET I OSLO
UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Building a typology of the linguistic functions of emoji:

A cross-linguistic and cross-platform corpus analysis of emoji in conversation

MASTER'S THESIS IN LINGUISTICS AND ITS APPLICATION FOR A MULTILINGUAL SOCIETY

Supervisors:

Hilde Hasselgård
Cédrick Fairon

2019 - 2020

Louis Escouflaire

2020

Building a typology of the linguistic functions of emoji: a cross-linguistic and cross-platform corpus analysis of emoji in conversation

Abstract

This thesis aims to identify and to define all the different linguistic functions that emoji can have in computer-mediated discourse. Emoji are graphical icons, representing attitudes or concepts, that became popular around 2010 and that can be used on any smartphone or computer to fill a variety of communicative roles inside a message or a conversation. A multimodal corpus, containing 1200 emoji and the messages in which they were used, was composed and analyzed in order to verify the effectiveness of an initial typology of emoji functions which was inspired by previous frameworks. The emoji in this corpus were collected on Facebook and on Twitter, in French and English messages, and in discussions about three different subjects. This allows for a comparable analysis of emoji functions depending on four external factors: platform, language, topic and speaker's gender. The quantitative analysis of the data, along with the qualitative analysis of specific examples of the corpus, led to the construction of a new typology of emoji functions, consisting of three primary functions (expressive, interpretative and referential) and five secondary functions (relational, politeness, emphatic, structural and aesthetic). Regarding variation, all four external factors appear to have a significant influence on the frequencies of the different emoji functions used in the corpus.

L'objectif de ce mémoire est d'identifier et de définir toutes les différentes fonctions linguistiques que les émojis peuvent avoir dans le discours médié par ordinateur. Les émojis sont des icônes graphiques, représentant des comportements ou des concepts, qui sont devenues populaires vers 2010 et qui peuvent être utilisées sur tout smartphone ou ordinateur pour remplir une variété de rôles communicatifs à l'intérieur d'un message ou d'une conversation. Un corpus multimodal, contenant 1200 émojis ainsi que les messages dans lesquels ils ont été utilisés, a été composé et analysé afin de vérifier l'efficacité d'une première typologie des fonctions des émojis, inspirée par des méthodologies antérieures. Les émojis dans ce corpus ont été récoltés sur Facebook et sur Twitter, dans des messages en français en en anglais, et dans des discussions traitant de trois sujets différents. Cela permet d'analyser comparativement les fonctions des émojis suivant quatre facteurs externes : la plate-forme, la langue, le thème et le genre du locuteur. L'analyse quantitative des données, ainsi que l'analyse qualitative de certains exemples du corpus, a mené à l'élaboration d'une nouvelle typologie des fonctions des émojis, composée de trois fonctions primaires (expressive, interprétative et référentielle) et de cinq fonctions secondaires (relationnelle, de politesse, emphatique, structurelle et esthétique). En ce qui concerne la variation, il apparait que les quatre facteurs externes ont une influence significative sur la fréquence des différentes fonctions des émojis dans le corpus.

Acknowledgments

First of all, I would like to thank my two supervisors. Cédrick Fairon, for supporting my project from the first day of this project and for always expressing his enthusiasm towards the very peculiar subject of emoji. Hilde Hasselgård, for providing me with the most valuable suggestions to improve my work, for reading and correcting each section of this thesis multiple times and for accompanying me every week during my year in Oslo.

I also want to thank the other teachers and academics who helped me during these two years. The members of the CENTAL in Louvain-la-Neuve, Thomas François, Hubert Naets, Anaïs Tack and Damien De Meyere, who found solutions to all the problems I encountered with corpus collection on social media. I am also very grateful to Serge Bibauw, who answered all my questions regarding statistical analysis in great detail.

Finally, I would like to thank my friends and my family for encouraging me every day and for allowing me to analyze each emoji they used for the past two years. My parents, for proofreading my thesis and for believing in my work even though they did not always understand everything. My girlfriend, for listening without complaint to all my wildest theories about emoji.

Thank you all 😊

Table of contents

Introduction

In everyday conversations, people smile all the time. They smile to express their joy; they smile to react to compliments. They smile when they apologize and they smile when they need to ask for help. They also smile to show that they enjoy the company of their interlocutors. And when they make a joke, they smile to make other people understand that they are being humoristic. But what happens when people are not able to see each other's smiles in a conversation, or any other facial expression? How can people discussing on the Internet understand each other properly if they lack all the information usually conveyed by non-verbal language? A potential answer to this question entered the daily lives of all Internet users more than a decade ago: emoji.

More than 60 billion emoji were sent every day in 2016, on Facebook only[1]. Between August 2019 and April 2020, around 20% of all Tweets contained at least one emoji[2]. As of March 2020, the Unicode Standard comprises 3,304 different types of emoji, which represent faces, animals, objects or some more abstract concepts. These numbers, among others, show that emoji have become popular linguistic devices used and understood by many people around the world. Yet, few researchers have considered emoji as an important matter in modern linguistics, even though they constitute a very interesting subject, and appear to be capable of accomplishing much more in conversation than it would seem at first.

Just like the paralinguistic cues in face-to-face conversation (i.e. facial expressions, intonation, laughter, etc.) that they mimic and the words that they replace, emoji are able to perform different communicative functions. For example, they can be used to show emotions and to signal irony, but also to manage clauses in an utterance and conversational turns in an interaction. However, all the different roles that emoji are actually able of playing in computer-mediated discourse have not been systematically identified and described yet. Therefore, the first objective of this research project is to build an operational typology of all the different functions of emoji in conversation. This new framework will provide a great tool for computer-mediated discourse analysis (CMDA) of interactions and contexts in which emoji appear. It may also have positive implications for technological purposes, for example through the improvement of "smart keyboards" which aim to predict and suggest the perfect emoji to add to a message.

[1] https://blog.emojipedia.org/5-billion-emojis-sent-daily-on-messenger/, last accessed on May 30, 2020.
[2] https://blog.emojipedia.org/emoji-use-in-the-new-normal/, last accessed on May 30, 2020.

In order to establish this new typology of the linguistic functions of emoji, 1200 emoji (and the messages in which they are used) will be collected in order to create a multimodal corpus. This corpus will be composed of messages in French and English, extracted from Twitter and Facebook, and discussing different topics. From there, the second objective of this study will be to perform a thorough investigation of emoji variation across different languages, social media platforms, topics and across gender, with a focus on emoji functions. This will give information on how the linguistic functions of emoji are influenced by linguistic, contextual and sociological factors, and will provide a testbed for whether our functional typology is operational for emoji in different languages and situations. Can emoji be considered as a universal form of communication, used and interpreted in the same way by each and every user of the Internet?

The first chapter of this thesis is dedicated to a detailed review of the literature that has already been produced on the subject. After explaining the importance of computer-mediated discourse and the role of social networking sites, the origin of emoji and their concept are explained. Then, an initial typology of the linguistic functions of emoji is presented, inspired by several previous frameworks and by personal observations. The state of the art ends with an examination of the different works produced by researchers on emoji variation according to gender, age, language and culture. The second chapter starts by detailing the two research questions that the thesis aims to answer. Then, the corpus used for the research project is described in-depth, from the collection of the data to its methods of analysis. The results of the investigation are presented in the third chapter. Various tables and statistics are used to visualize these results, which are supported by many examples from the corpus. Finally, the fourth chapter discusses in great detail the findings of the analysis. The final typology of emoji functions is presented and both research questions are answered, followed by some final observations on the limitations and strengths of the study.

1. State of the art

This chapter aims to provide a general background of the different subjects that will be addressed in the thesis. It is divided in five sections, each of which discusses and puts into perspective the findings of prior studies. Some points of this literature review are supported by examples, the majority of which were made up for the purpose of this overview.

1.1. Computer-mediated discourse analysis (CMDA) and social media

1.1.1. The rise and evolution of CMDA

Language has always evolved along with society. Big sociological phenomena such as human migrations and regular exchanges between different groups of people have led to major linguistic changes across the ages. Economic and technological factors are also very important in the evolution of any language. For example, the invention of printing and the Gutenberg revolution marked the coming of a new era characterized by a more democratic access to literacy and the emergence of modern linguistic features, including fixed rules of punctuation and the standardization of grammar and spelling in many languages.

In quite a similar way, when the World Wide Web was rendered public in the mid-1990s, some linguists had already pointed out the effects that the use of computers could have on the English language (Baron, 1984; Herring, 1996). As the Internet grew in popularity, and as more and more people from different backgrounds and speaking different languages started using it in their work and in their everyday life, computer-mediated communication (CMC) attracted the attention of authors from several fields of research (Herring, 2019). While sociolinguists were investigating the reasons behind code-switching on discussion forums, lexicologists tried to identify the plethora of neologisms popping up in these new genres of communication (Androutsopoulos, 2006).

Never in history had people been able to write and be read as freely as through the Internet. Gutenberg allowed everyone to become a reader; the Web allowed everyone to become a writer. Of course, this process of democratization (which has yet to be fully achieved) gave birth to new ways of writing (McCulloch, 2019). In order to refer to those developing linguistic characteristics, journalists and researchers came up with several designations, like “Internet-slang” or “Netspeak”. However, one name that most linguists agree on is “computer-mediated discourse”, commonly abbreviated as “CMD” (Herring, 2019).

Although it would have been unthinkable forty years ago for researchers to analyze discourse produced on the Internet with the same techniques as they analyzed "offline" language, the tables have now turned (Panckhurst, 2006). Over the last three decades, hundreds of academic papers have been written using techniques of computer-mediated discourse analysis (hereafter CMDA) to examine samples and corpora of emails, instant messages and blog posts. Those studies have been carried out using essentially the methods of traditional discourse analysis applied to CMD (Herring & Androutsopoulos, 2007). But according to American linguist Susan Herring, who has been studying CMD ever since 1992, new forms of communication have appeared in the last two decades, suggesting that the entire concepts of CMD and CMDA need to be reshaped. With the rise of "interactive multimodal platforms" such as video conferencing systems (e.g. Skype, Microsoft Teams, Omegle) and video-sharing platforms on which people can answer not only with text but also with voice and image (e.g. Facebook Messenger, Instagram), it has become obvious that text is no more the only mode of communication in CMD (Herring, 2019). This is even more evident on social media websites such as Facebook or Twitter, where users reply to each other's messages with memes, animated GIFs and emoji. Those features, which Herring refers to as "graphicons", are part of a large ensemble of linguistic characteristics more or less specific to CMD, and that have been detailed in previous research.

1.1.2. The linguistic features of CMD

Even though non-textual modes of communication have been gaining a lot of attention in CMD in recent years, written language was for a long time the only way to convey information from one user to another on the Internet. Of course, this is not a problem when it comes to formal language, as the writer is expected to avoid the expression of emotions, for example in academic writing, on serious websites or in formal emails. However, in other contexts, such as in instant messaging and forum discussions, informal conversations are prevalent. Therefore, CMD users have always tried (and succeeded) to find ways to adjust their writing styles accordingly to the context of the conversation.

The presence of various marks of orality in discourse is certainly the most significant characteristic of CMD. In oral conversation, non-verbal signs play a very important role and can add whole new layers of signification to an utterance. Intonation, gesture and laughter, among other cues, all contribute to building the meaning of a message (Troiano & Nante, 2018). In order to overcome the lack of these paraverbal signs, which have been proven to be essential

to oral communication (Hall et al., 2019), several alternatives have appeared. Writing words or entire messages in uppercase letters, as in example 1 below, can be used as a way to convey anger, fear, or to simulate a loud (or even shouting) tone of voice (Marcoccia, 2000). Other ways to reproduce intonation in CMD include the repetition of letters or punctuation marks in order to express emphasis (Maíz-Arévalo, 2015), as in example 2, or the use of emoticons, which are combinations of punctuation signs representing human faces and used to mimic facial expressions and gestures, and the use of emoji. This particular feature will be examined in more detail in section 1.2. Moreover, onomatopoeias such as *pfff, haha* or *ummm* are commonly used in CMC conversations to communicate the same reactions or feelings as their face-to-face equivalents: in this case irritation, laughter or hesitation (Petitjean & Morel, 2017).

(1) WHY WONT YOU ANSWER

(2) im *soooo* down for *thisss*

Another key linguistic feature of CMD is the use of new lexical items. First, a large number of neologisms have naturally been created by CMC users to refer to new concepts brought with the rise of computers and the Internet. Crowdsourced online dictionaries such as the Urban Dictionary[3] and the Wiktionary[4] give an overview of all these new nouns, verbs or phrases. Some of those words, like *hashtag* or *spamming*, have even entered the English spoken language. But one specific property of CMC precipitated many changes in the lexical field: the size of message buffer (Herring, 2007). On various forums or social networks, the amount of characters that can be used in one single message is limited. Up to 2017 on Twitter, users had to carefully pick their words in order to fit their whole statement in 140 characters (280 since 2017). Back when most telephone companies made their clients pay a fixed price per text message sent, users needed to find a way to communicate efficiently without crossing the size limit (i.e. 160 characters, usually). This economic factor led to the creation of a variety of abbreviations and acronyms, along with the use of simplified grammar and spelling (Herring, 2019). Messages like the one in example 3 were quite common in the early 2000s. A return to more classic spelling has been observed over the past fifteen years, though some fixed forms of abbreviated words and acronyms still appear to be used widely by CMD users. One of those

[3] https://www.urbandictionary.com/, last accessed on 30 May, 2020.
[4] https://en.wiktionary.org/wiki/Wiktionary:Main_Page, last accessed on 30 May, 2020.

is the world-famous word *lol* (or *LOL*), which occupies a very particular spot in the history of CMDA.

(3) how r u? hop u'll b busy in ur stdy.Ma 2. When r v gona meet 4 dat work?
(How are you? I hope you will be busy in your study. Me too. When are we going to meet for that work?)
(Rafi, 2008)

Lol literally means "laughing out loud" and can be used in a message to show that the speaker is actually laughing, or simply amused by the previous message or by the general context of the conversation (Tagliamonte & Denis, 2008). Linguists later realized that *lol* can have different functions than just displaying the writer's emotion: it can also signal irony in a message, act as a linguistic hedge or even as a discourse marker. The use of *lol* extended to languages outside English, often with the same functions, as Deniz Uygur-Distexhe (2007) showed with French CMD examples. However, *lol* has slowly lost its popularity and has been replaced by emoji and by onomatopoeias reproducing laughter, such as *haha* or *hehe* (Adamic et al., 2015; Cougnon, 2015), which are even taking over its linguistic functions (Escouflaire, 2018b). Nowadays, it seems that *lol* is primarily being used in an ironic way, to mean that something is "not that fun" (McCulloch, 2019).

1.1.3. CMD genre classification

It is important to note that not all linguistic features apply to all the various genres of CMD. Research has shown that some users adopt different writing styles when they are writing on online forums or on instant messaging apps (Androutsopoulos, 2006). This has led linguists to try to classify those CMD subtypes into a systematic scheme, in order to get a better insight into each of them and to compare them more efficiently. Classification is even more important for CMD corpora studies, in which data from different websites or apps is investigated (Chanier et al., 2014).

Susan Herring developed a popular classification scheme of what she refers to as "socio-technical modes" (Herring, 2007). Two types of criteria can help cataloguing the different modes: medium factors and social factors. First, the medium factors (M1 to M10), listed in Table 1, are the elements that are influenced by the device, website or app which users are writing and communicating with. Ten different criteria have been proven to have an influence

on the users' discourse. First, synchronicity (M1) relies on the simultaneous (synchronous CMC) or non-simultaneous presence (asynchronous CMC) of participants in a conversation (Chan, 2011). For example, in chatrooms, users reply to each other almost at the same pace as in real-time conversation, whereas emailing someone does not require that person to be logged in at the same time. Then, message transmission (M2) relates to whether the messages are sent line by line (e.g. on some instant messaging apps) or message by message (e.g. on forums). Some new mediums, like most instant messaging apps (e.g. Facebook Messenger, Viber, WhatsApp), allow their users to know if another user is currently typing a message or if a given message has been seen by the receiver, a feature that can play an important role in conversation (Hoyle et al., 2017). Persistence of transcript (M3) refers to the amount of time after which a message is erased from the conversation. Emails and forum posts are archived virtually forever (until they are potentially deleted by the receiver or the sender), but messages on "ephemeral conversation" apps like Snapchat disappear after a short moment (Cavalcanti et al., 2017). Size of message buffer (M4) is the character limit of a message, already mentioned in section 1.1.2. The channels of communication (M5) concern the different types of communication available: is text the only resource, or can users send pictures, graphicons (e.g. pictures, GIFs, emoji) or communicate through audio and video? Anonymous messaging (M6) refers to the ability to hide one's name or use pseudonyms, and private messaging (M7) to apps allowing non-publicly shared conversation. Websites or apps with filtering (M8) allow users to ignore messages from other specific users, and those providing a quoting feature (M9) let users cite previous messages from the conversation in their own messages. Message format (M10) refers to how conversations are visually displayed and in what order the messages are presented.

Table 1. Medium factors for CMD genre classification (Herring, 2007).

M1	Synchronicity
M2	Message transmission
M3	Persistence of transcript
M4	Size of message buffer
M5	Channels of communication
M6	Anonymous messaging
M7	Private messaging
M8	Filtering
M9	Quoting
M10	Message format

Secondly, eight social factors have been identified by Herring (2007), in a scheme inspired by Dell Hymes' (1974) "SPEAKING" model. Participation structure (S1) covers the number of participants in a CMC conversation, whether it is public or private, and how much each speaker engages in the exchange. Participant characteristics (S2) concerns information about the gender, age, sociological background and proficiency (with the language in use and with CMD) of the users, their relationships, and all the other individual details that can be helpful for the analysis. Purpose (S3) refers to the goal of the conversation (e.g. professional, social, recreational) and Topic (S4) to the theme around which the interaction revolves. The Tone (S5) of a conversation is the general mood in which it occurs (e.g. serious, flirty, hostile). Then, Activity (S6) concerns the general context of the interaction (e.g. job interview, game, phatic conversation). The Norms (S7) are the rules defined by the app on which the CMD exchange occurs, for example: whether insults are permitted, how new users enter the conversation, and which linguistic features are commonly employed by the users. Finally, Code (S8) designates the language and register of language used for the interaction.

Table 2. Social factors for CMD genre classification (Herring, 2007)

S1	Participation structure
S2	Participant characteristics
S3	Purpose
S4	Topic
S5	Tone
S6	Activity
S7	Norms
S8	Code

Applying this classification scheme to CMDA allows researchers to analyze CMD data very precisely and to determine with accuracy how to compare samples gathered from different websites, apps or devices. Herring's model can not only be used to identify differences between various socio-technical modes, but also between seemingly very similar websites: two social networking sites, for example. However, it should be noted that this model was built in 2007, and Herring herself admits that the Internet and CMD have considerably evolved since then (Herring, 2019). This model should therefore be used with caution, as will be explained in more detail in section 2.2.2.

1.1.4. Social networking sites

At the start of 2019, 2.32 billion people worldwide were monthly users of Facebook, the most popular social media in history (Wilberdin & Wells, 2019). Social media can be defined as "Internet-based channels" providing an interpersonal communication network on which users are given the ability to consume information, produce it themselves and interact with each other freely (Carr & Hayes, 2015). This definition includes several types of websites, from forums and blogging platforms such as Wordpress to networks like Twitter and Instagram. The latter are part of a subcategory of social media: social networking sites (Page et al., 2014).

The essential features of social networking sites (SNSs) are the existence of personal profiles containing information about each user and the possibility for users to manage their connections with other users efficiently. Participants are able to share publicly (or semi-publicly) various types of content (Ellison & Boyd, 2013). The popularity of SNSs has increased enormously over the last decade and they have become an important part of millions of people's daily lives, giving them the chance to engage in social exchanges with family members, acquaintances and even strangers from any place in the world, at any time of the day (Lin & Qiu, 2013). People use SNSs for various purposes: social interaction, entertainment, self-expression, convenience, or simply for passing time (Alhabash & Ma, 2017). Research has shown that daily consumption of SNSs has positive effects, such as building and maintaining relationships, as well as negative consequences, including voyeurism and exhibitionism (Mäntymäki & Islam, 2016).

There are hundreds of different SNSs, but only a handful of them are internationally popular. For example, even though Weibo, Qzone and WeChat all count more than 300 million monthly users, they are almost exclusively used by Chinese speakers. In the English-speaking world and in most other countries, Facebook is by far the most popular SNS, followed by Youtube (1.9 billion monthly users), WhatsApp (1.5 billion) and Instagram (1 billion). Then, Tumblr, Tiktok, Reddit, Twitter and Snapchat all have between 300 and 500 million monthly users (Ortiz-Ospina, 2019).

Analyzing discourse on social networking sites is very interesting, as they allow people from very different backgrounds to write about any subject they want, and how they want. However, users do not usually behave in the same way on different SNSs, mainly because they do not use them for the same purposes. For example, in 2019, Facebook was used by college students primarily for convenience, whereas Twitter is mostly seen as a source of entertainment by the same users (Alhabash & Ma, 2017). This leads CMD users to use different writing styles

depending on the SNS on which they are communicating. Lin & Qiu (2013) discovered that Facebook posts tend to show greater verbal immediacy (i.e. shorter words and more frequent use of first person singular) than Tweets, thus reflecting through language the fact that Facebook mostly builds on real-life relationships, while Twitter interlocutors do not usually know each other. Moreover, it was found that Facebook users were more likely to express emotions through the use of non-verbal cues, such as emoticons, whereas Tweets contained more affective words (Lin & Qiu, 2013). This leads to the hypothesis that there are linguistic differences between messages produced on different SNSs.

1.2. What are emoji?

1.2.1. Historical overview

On 19 September 1982, computer scientist Scott Fahlman noticed that the computer messaging system through which professors of Carnegie Mellon University communicated lacked something important: a way to signal if a message was intended as a joke or not. Fahlman proposed two strings of three characters that would join every message in need of disambiguation. The string ":-)" indicated that the message was a joke or contained irony, and ":-(" that the message was to be interpreted as serious (Dresner & Herring, 2010). These "punctuation faces" rapidly became popular, expanded outside the university, and developed into several variants, such as ";-)", ":-P" or ":'-(". They were named "emoticons", a portmanteau word of "emotion" and "icon", because they began to be used as a way for the writer of a message to express his emotional state when sending the message (McCulloch, 2019). A few years later, in 1986, appeared in Japan the first "kaomoji", from Japanese *kao* ("face") and *moji* ("character"), strings of characters very similar to emoticons, except that they were not to be read sideways: "^_^" or "T_T", for example (Markman & Oshima, 2007).

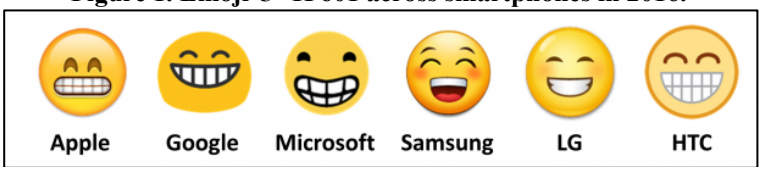
Although emoticons remained very popular in Western CMC until the late 2000s, an alternative to kaomoji was introduced in Japan in 1997 by Shigetaka Kurita, a designer of Japanese mobile operator DoCoMo (Lebduska, 2014). In order to appeal to Japanese teenagers, Kurita invented the "emoji", from *e* ("picture") and *moji* ("character"), small colorful graphics available in 176 different shapes, representing human faces, animals and objects. Their success was immediate, and soon every mobile phone company in Japan integrated emoji into their system (Blagdon, 2013). Ten years later, when Apple's first iPhone was launched worldwide

in 2007, Japanese users requested the ability to use emoji directly from their iPhone's keyboard (Tuttle, 2016). However, in order to allow users with different phones to have text conversations containing the same emoji, these new symbols had to be approved by the Unicode Consortium.

The Unicode Consortium is a small committee of people working for major hardware companies all over the world who decide which characters will be added to the Unicode standard – the international "character list" implanted in every CMC device. Every character is assigned a specific code, which allows it to be understood in the same way by two or more different devices. For example, the Latin capital letter "A" is encoded as "U+0041" in the standard, and the Copyright symbol "©" is encoded as "U+00A9". However, in 2007, when a particular emoji was sent from an iPhone to a Samsung mobile phone for example, it was very likely that the emoji would not be understood correctly by the receiving device and that the reader would see an amorphous grey square appear instead of the intended emoji. Therefore, in 2010, the Unicode Consortium considered that emoji were not just a passing trend, and 860 different emoji were added to the Unicode Standard (McCulloch, 2019).

Following this "emoji revolution", CMC users outside Japan discovered the symbols and adopted them quite rapidly. In 2015, the "face with tears of joy" emoji, "😂" (U+1F602), was used more frequently on social media than the most popular emoticon, ":)", proving that these new single-character symbols were becoming way more popular than their multiple-characters ancestors (Medlock & McCulloch, 2016). During the same year, emoji were found in more than 2% of all Tweets (Novak et al., 2015) and in 50% of all messages posted on Instagram (Dimson, 2015). Since 2010, the Unicode Consortium has been making new updates by adding new emoji. By the time of the last update, in March 2020, there were 3304 different types of emoji, according to the website Emojipedia[5].

Figure 1. Emoji U+1F601 across smartphones in 2016.



Although the Unicode standard is universal, it is important to note that the graphical rendering of each individual character still depends on the device on which it is received. Apple, Android, Samsung, Facebook and Twitter, among others, all use the same code for the "beaming face with smiling eyes" emoji (U+1F601), for example, but it is not rendered exactly

[5] https://emojipedia.org/faq/#how-many, last accessed on 30 May, 2020.

identical from one platform to another. These slight differences can cause misinterpretation in messages sent across different devices (Miller et al., 2016). In 2017, Apple was pressured by its users to modify the appearance of the aforementioned emoji, because it was too different from the other platforms' renderings and caused many misunderstandings between users of different devices (Miller et al., 2018). Figure 1 above illustrates the past differences (in 2016) between the designs of this particular emoji.

1.2.2. Defining emoji

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). The popularity of emoji seems to reflect the attraction of our modern digital society for visual cues. But emoji are not the first graphical representations to appear in text. Smiley faces were found in written postcards long before the invention of emoticons. Lewis Carroll also used to send "rebus letters" filled with drawings used to replace some words in his sentences (McCulloch, 2019). Some linguists even discuss the intriguing similarity between emoji and pictographic systems of communication, such as ancient Egyptian hieroglyphs (Brisson, 2015; Alshenqeeti, 2016). However, Oxford Dictionaries' decision to elect an emoji as their "word of the year" raises an important question: can emoji really be considered as words?

(1) *emoji*: "A digital image that is added to a message in electronic communication in order to express a particular idea or feeling." (Cambridge Dictionary)[6]
(2) *emoji*: "A small digital image or icon used to express an idea, emotion, etc." (Oxford Dictionaries)[7]
(3) *emoji*: "Any of various small images, symbols, or icons used in text fields in electronic communication (as in text messages, e-mail, and social media) to express the emotional attitude of the writer, convey information succinctly, communicate a message playfully without using words, etc." (Merriam-Webster)[8]

None of these three definitions, extracted from renowned English dictionaries, characterize emoji as "words". Instead, they are jointly described as "small digital images or icons used in CMC to express an idea or emotion." Although they refer to items and concepts,

[6] https://dictionary.cambridge.org/dictionary/english/emoji, last accessed on 30 May, 2020.
[7] https://www.lexico.com/en/definition/emoji, last accessed on 30 May, 2020.
[8] https://www.merriam-webster.com/dictionary/emoji, last accessed on 30 May, 2020.

just like words do, the relationship with their referents is iconic, not arbitrary. This feature makes it very difficult – and sometimes impossible – to build meaningful sentences using only emoji, as it was experienced by Fred Benenson, who attempted to translate Herman Melville's Moby Dick entirely into "emoji language" (Benenson, 2010; Radford et al., 2016). Emoji lack ways to represent a lot of abstract concepts, actions or relations, which can though be denoted with textual nouns, verbs or conjunctions. Deictic words (*today*, *here*, *this* house), for example, are non-existent in emoji language: it is easy to use the "dog" emoji, "🐕", in order to refer to the concept *dog*, but it would be very difficult to refer to "*this particular* dog" or to "*my* dog" without the use of actual words. Moreover, metalanguage is not possible with emoji, as they appear to be unable to refer to themselves. Hamza Alshenqeeti (2016) considers emoji as a form of paralanguage that only functions when it is associated with a more efficient code (e.g. the English language). Emoji thus seem to work a lot better as a supplement to language than as a language itself.

In face-to-face speech, features such as prosody, gesture or laughter are able to reveal the emotions of the speaker, add information to the message or even modify its meaning. Emoji, and before them emoticons, have been regarded by several authors as ways to overcome the lack of these nonverbal cues in CMC (Marcoccia & Gauducheau, 2007; Troiano & Nante, 2018). Linguists Gretchen McCulloch and Lauren Gawne (2018), while looking at the common use of repeated emoji (two or more identical emoji used in a row), proposed to build a parallel with "beat gestures". A beat gesture is a gesture that contains "a repetitive up-down or side-to-side rhythm", such as clapping hands or putting your thumb up while moving your hand back and forth. Gestures accompany speech utterances, and beat gestures emphasize the meaning of these utterances. Emoji – especially those which represent human faces or body attitudes – can thus be seen as digital gestures that complement the meaning of a message, while adding emphasis when repeated. The quite unusual use of sequences of different emoji (in comparison with the very high recurrence of identical emoji sequences) constitutes another argument that emoji do not work like words (McCulloch & Gawne, 2018).

Unlike emoji, emoticons are symbols built of strings of punctuation signs. Nowadays, many devices and platforms automatically convert some emoticon sequences to the corresponding single-character emoji (Schneebeli, 2017). For example, if a user types ";-)" on an iPhone or on Facebook, the "winking face" emoji, "😉", will automatically replace the emoticon. Yet this is not the only link between punctuation and emoji: Dresner and Herring (2010) observe that emoticons can be compared to question marks and exclamation marks,

because in the same way they work as an "expansion of text". They give an insight into the mood or the tone of the message. Moreover, emoji are very often found in final position in sentences and in whole messages, and they have been proven to be a popular alternative to the full stop (".") in CMD (Sampietro, 2016). Using a full stop to end a sentence can be interpreted by some users of instant messaging apps or as an indicator of negative emotion, or even as rude (in contrast with the presence of emoji at the end of the sentence), which can lead to misinterpretations between different users (Baron & Ling, 2011; Escouflaire, 2018a). However, emoji display a wider range of functions in conversation than just those of punctuation (Schneebeli, 2017), which makes it hard to categorize them as punctuation signs. We will further look at the functions of emoji in section 1.3.

Another linguistic category in which emoji could possibly fit is the category of discourse markers. According to Deborah Schiffrin (1987), a discourse marker is a word or a group of words that helps managing the flow and organization of discourse. In the same way as discourse markers, emoji are independent from the syntax of the sentences and can be used to structure several ideas inside a single message or between different messages (Sampietro, 2016). Authors have shown that e-laughter, similarly to laughter in real-life conversations, can be a productive tool for managing CMD conversations: words and onomatopoeias such as *lol* and *haha* can work as discourse markers (Uygur-Distexhe, 2012; Petitjean & Morel, 2017; Escouflaire, 2018b). Then, a study on Facebook CMD in 2015 showed that emoji were becoming the most popular way to express e-laughter (Adamic et al., 2015; Cougnon, 2015). Considering that many uses of emoji overlap with the functions of discourse markers identified by Crible (2017), the categorization of emoji as discourse markers could be a path worth to investigate further.

Overall, emoji do not seem to fit perfectly in one of the four categories mentioned above (words, beat gestures, punctuation signs or discourse markers), mostly because none of these classes encompass the multiple functions of emoji. Therefore, Herring and Dainas (2017) introduced a new term: *graphicon*, a blend of *graphical* and *icons*. Graphicons include not only emoji, but also emoticons and kaomoji, as well as other graphical means of communication with conversational uses in CMD, such as GIFs, images and videos. Following this new classification, emoji can be defined as follows:

Graphical icons, representing attitudes or concepts, used in computer-mediated discourse in order to fill a variety of communicative roles inside a message or a conversation.

1.2.3. The different types of emoji

As of May 2020, a total of 3304 emoji are officially listed in the Unicode Standard.[9] Most digital keyboards on smartphones and apps divide these emoji into eight different lists, making them easier to find for their users (Pohl et al., 2016). Emojipedia, a website created in 2016 by Jeremy Burge and considered as a reference on emoji by many scholars (Barbieri et al., 2016a; Miller et al., 2016, McCulloch, 2019), divides its articles on the basis of the same eight emoji categories: "Smileys & People", "Animals & Nature", "Food & Drink", "Activity", "Travel & Places", "Objects", "Symbols" and "Flags". Some CMD devices and platforms also include a list gathering their user's most commonly used emoji, as well as a list of their most recently used emoji.

In October 2019, the Unicode Consortium published a table containing all the different types of emoji ranked from most to least frequent, across multiple platforms. They claimed that frequency was "one of many considerations taken into account when deciding on which emoji to add in the next Unicode update" (Unicode, 2019). These statistics revealed that the "face with tears of joy" emoji "😂" (the aforementioned "word of the year 2015") is still, four years later, the most used emoji. The top five list is completed by the "red heart" emoji ("❤️"), the "smiling face with heart-eyes emoji" ("😍"), the "rolling on the floor laughing" emoji ("🤣") and the "smiling face with smiling eyes emoji" ("😊"). The whole "top 20" is presented in Table 3 below.

Table 3. The 20 most frequently used emoji as of October 2019

1	😂	11	😅
2	❤️	12	👏
3	😍	13	😁
4	🤣	14	♥️
5	😊	15	🔥
6	🙏	16	💔
7	💕	17	💖
8	😭	18	💙
9	😘	19	😢
10	👍	20	🤔

[9] https://emojipedia.org/stats/, last accessed on May 30, 2020.

According to Ilaria Moschini's (2016) socio-semiotic analysis, the international popularity of the "face with tears of joy" emoji is due to the fact that it combines the exaggerated graphical emotion of the Japanese manga culture and the optimistic commercial *ethos* of the American mindset. It has also been classified as one of the emoji with the least potential for misinterpretation between speakers (Miller et al., 2016). Described by Emojipedia as "widely used to show something is funny or pleasing"[10], the "face with tears of joy" emoji seems to share the same purpose as e-laughter words (like *haha* and *lol*); the common use of this emoji for expressing in a very straightforward way that something is humorous could explain its success. In Unicode's top five, another emoji graphically represents extreme hilarity, the "rolling on the floor laughing" emoji. This can be compared to the exaggeration found in acronyms that were really popular before the rise of emoji, such as *lmao* ("laughing my ass off"), *rofl* ("rolling on the floor laughing") or even *lol* ("laughing out loud"). Those words were used hyperbolically by speakers who were probably not laughing that much behind their screens (McCulloch, 2019).

From the eight categories, it seems clear that the "Smileys & People" emoji are the most popular. Inside this first category, three subclasses can be distinguished: "facial emoji", "people emoji" and "gesture emoji". Facial emoji (e.g. "😊" or "😍") are the most common of all, and also the most stereotypical. They derive directly from emoticons and kaomoji and are usually assigned to the most basic functions of emoji (i.e. the expressive and interpretative functions, as detailed in section 1.3). According to the Unicode Consortium' statistics, ten of the twenty most used emoji are facial emoji (Unicode, 2019). From these ten facial emoji, only two express negative emotions ("😭" and "😢"). This correlates with Lin and Qiu's (2013) study, which showed that CMD users are a lot more likely to display positive than negative emotions, especially on public social networking sites. Then, people emoji represent various types of bodily attitudes, job roles, couples and families. Just like gesture emoji, which are graphical renderings of hand or body gestures (the most common are the "folded hands" and the "thumbs up" emoji), people emoji can be adjusted with "emoji modifiers": skin tone and gender. These two features were added in 2015 to the Unicode standard with an inclusive purpose, even though they also led to many controversies and to the apparition of "emoji stereotypes" (Stark & Crawford, 2015; Barbieri & Camacho-Collados, 2018).

Only seven of the twenty most used emoji are not facial, people or gesture emoji. Six of those are hearts, coming in different shapes and colors, intact or whole. Hearts, although they

[10] https://emojipedia.org/face-with-tears-of-joy/, last accessed on May 30, 2020.

are part of the “Symbols” category, are more popular than many facial emoji. As intimacy is one of the main reasons why users claim to use emoji (Wiseman & Gould, 2018), it is not surprising that such a universal symbol of love is so successful.

The other six categories mostly contain emoji that iconically represent tangible items, activities or locations. Their primary function is referential (see section 1.3.7.) and their meaning is almost always that of the concept they represent. For example, the “dolphin” emoji has the same meaning as the word “dolphin”, and the “trumpet” emoji refers to nothing more than the concept of a brass wind instrument. However, there are some exceptions. The “fire” emoji, “🔥”, is the only character from these six categories to appear in the twenty most-used emoji (Unicode, 2019). The obvious reason for this popularity is that the main meaning of the “fire” emoji is not purely referential: it is mostly used to express excitement rather than actual flames (Gawne & McCulloch, 2019). CMC users have assigned derivative meanings to many emoji, exceeding their original referential purposes. The “eggplant” emoji, “🍆”, for example, is often used euphemistically to add a sexual meaning to a message, based on the emoji’s resemblance with male genitalia (Weissman, 2019). The “peach” emoji, “🍑” gained the same connotation, but also acquired political meaning in September 2019, when it was used as a protest symbol against Donald Trump’s impeachment inquiry (Zanotti, 2019). Aside from that, emoji have also been used with communitarian meanings, such as the “unicorn” emoji, “🦄”, used by the LGBTQ+ community as an identification symbol (Quito, 2019).

These examples show that, most of the time, the meaning of emoji heavily relies on the context of the message or the conversation in which the emoji appear. One single emoji can be used to express different things depending on the way in which the writer uses it. These potential meanings often depend on the functions in which emoji are used. The various functions of emoji are exposed in the following section of this chapter.

1.3. The various linguistic functions of emoji

Though it may seem at first sight that the only reasons why people use emoji in CMD are to transcribe physical emotions and to represent items and concepts iconically, it has been demonstrated by several linguists that the roles of emoji are much more complex. Emoji appear to be able to perform a wide variety of linguistic functions that users consciously or unconsciously apply and understand. This chapter presents an overview of the different emoji functions that were identified by previous searchers. More precisely, as no academic consensus on a classification scheme for emoji functions has yet been found, this section aims to provide a synthetic compilation of the most prominent typologies proposed since the early 2000s.

Indeed, some of the papers cited in this section focus on emoticons and their functions, because emoji only grew popular worldwide after the year 2010 (as explained in section 1.2.). However, Herring and Dainas (2017) state that emoji fulfill almost exactly similar roles as emoticons, at least regarding facial emoji and hearts (as those are among the only concepts represented by emoticons). Although the visual and practical differences between emoticons and emoji could lead to slight functional disparities, this has not been observed by any researcher. For this reason, studies investigating emoticon functions were included among those about emoji functions. Even though it was inspired by previous frameworks, the typology of emoji functions presented in this section is new and was built for the purpose of this thesis.

Our framework distinguishes between seven different emoji functions, plus a few minor functions that could be considered as "nonlinguistic". While some of the papers that inspired this methodology found up to nine different emoji functions (Herring & Dainas, 2017; Beißwenger & Pappert, 2019), other identified only three or four functions (Cramer et al., 2016; Tang & Hew, 2019). In all of these previous classification schemes however, the expressive function is almost always considered as the primary function of emoji in CMC.

1.3.1. The expressive function

One of the most natural and obvious functions of emoji, and before them 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 speakers 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 “smiling face” emoji, anger with an “angry face” emoji or support with a “raised fist” emoji, for example (Schneebeli, 2017).

There are two different ways in which emotions can be expressed with the use of emoji in CMD. First, expressive emoji are able to indicate the emotional state of the speaker when sending the message (Marcoccia & Gauducheau, 2007). They work as “indicators of affective states” (Dresner & Herring, 2010; Na’aman et al., 2017), as in example 4 below. 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 a speaker stands in a conversation (Schneebeli, 2017). In the same way, they can be used to react to a previous message in a conversation, or to a situation, as in example 5. In Facebook comment threads, for instance, emoji are often found on their own (without verbal content), used as a reaction to a previous message in the thread (Herring & Dainas, 2017). Beißwenger and Pappert (2019) call this use the “evaluative or commentary function” of emoji.

(4) *I’m going to a party tonight* 😄

(5) A: *What did you think of the main character’s death?*
B: 😢😢😢

Quite logically, facial emoji are the main category of emoji to which the expressive function can be assigned, as they represent directly how the speaker’s emotions would be materialized 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 speaker, while the “praying hands” emoji can express admiration. The “red heart” emoji (along with its many variants) is also used by CMd speakers to express love or a strong positive feeling towards something (Schneebeli, 2017).

1.3.2. 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, interpretative emoji replace these features and work as indications to how a sent message should be understood by the receiver (Herring & Dainas, 2017). 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 message. These emoji guide the interpretation of a message, and by doing this they can help clarify the intention of the speaker and eliminate potential misunderstandings between interlocutors (Amaghlobeli, 2012). Sentences 6 and 7 below show how the presence of an emoji can influence the interpretation of a message: the "grinning face with smiling eyes" emoji suggests that the speaker is excited about eating pizza again, whereas the same sentence with the "crying face" emoji at the end guides the receiver towards a negative interpretation of the same message. Spina (2018) refers to such emoji as pragmatic markers, as they allow users to "infer contextually appropriate meanings in text", in the same way as the "contextualization cues" in oral discourse, as theorized by Gumperz (1982).

(6) *Mom said we're eating pizza again today* 😄

(7) *Mom said we're eating pizza again today* 😢

A context in which emoji are often used with their interpretative function is that of irony and sarcasm. Paul Grice (1975) defined irony as a blatant violation (flouting) of his first maxim of quality (*Do not say what you believe to be false*). In an ironic message, a speaker says or writes the opposite of what he actually means. In CMD, emoji can be used as effective markers of irony, in order to let the receiver know that the 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 CMC (González-Ibáñez et al., 2011; Weissman & Tanner, 2018).

Dresner and 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 looked into emoji as "illocutionary force markers", as indicators of the intended action of a message. 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 8, for instance, the "face with tears of joy" emoji indicates that the statement should not be taken literally, because the speaker is being ironic.

(8) *I hate you so much* 😂

1.3.3. The relational function

Emoji can also be used to maintain and manage relationships between CMD speakers. Like eye-contact or laughter in face-to-face conversation, they can have a phatic role and work as indicators of closeness (Marcoccia & Gauducheau, 2007) or markers of familiarity between users (Spina, 2018). Messages accompanied by emoji are perceived as more pleasant and as displaying a higher level of intimacy (Janssen et al., 2014; Tang & Hew, 2019). But even though they can be a very efficient tool for increasing intimacy in digital relationships, research has shown that relational emoji used in already intimate relationships (e.g. between romantic partners) do not have the same effect (Rodrigues et al., 2017).

Then, they can also work as a way to request familiarity from the receiver, that he or she can accept by using emoji in return. 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 speaker (Spina, 2018). Speakers who use emoji tend to be considered as "more extravert and agreeable" than those who do not include emoji in their messages (Fullwood et al., 2015).

(9) A: *The exam is at 10 am tomorrow!*
B: *Thanks!!*
A: *No problem* 😊
B: 😊

Emoji can also have an "economizing" phatic function, when they are used as means of maintaining a "conversational connection" between speakers 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 speaker 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 9 above illustrates such a conversation: speaker B ends the conversation with a single relational emoji, acknowledging speaker A's second turn. Speaker 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".

1.3.4. The politeness function

Another function of emoji concerns their capacity to work as linguistic hedges, according to the politeness theory introduced by Brown and Levinson (1987). Inspired by the concept of *face* introduced by Goffman (1967), they analyzed the conversational strategies used by speakers to weaken potential face-threatening speech acts (e.g. request, order, criticism, disagreement, accusation). One of these strategies is the addition of lexical softeners, or linguistic hedges, in a message: words and expressions, such as *please* or *maybe*, that mitigate the threatening value of a speech act. Marcoccia and Gauducheau (2007) identified emoticons used with the same purpose as these linguistic hedges and called them "politeness emoji".

Several authors have since studied the use of emoji as linguistic hedges in CMD. Skovholt et al. (2014) found emoji used to soften various types of face-threatening acts in a corpus of workplace emails, while Alden (2019) investigated their presence in online discussion forums. In example 10 below, the "smiling face with smiling eyes" emoji is used to weaken a face-threatening order, whereas in sentence 11 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 relationships the simple absence of emoji in a message can make a speech act feel more menacing for the receiver (Baron & Ling, 2011).

(10) *I'm here, open the door* 😊

(11) *You shouldn't use so many abbreviations in the first paragraph* 😉

While the politeness function is acknowledged by most classifications of emoji functions, many researchers 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 avoid ambiguity, the typology presented in this chapter follows other papers in which politeness is considered as a separate emoji function (Marcoccia & Gauducheau, 2007; Beißwenger & Pappert, 2019).

1.3.5. The emphatic function

There are two different manners in which emoji can have an emphatic function. First, emoji can emphasize an emotion or a stance that has already been expressed through the verbal content of the message. By "mirroring" the emotional content of the message in a redundant way, they reinforce its value (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 soften the message even more, as in example 12. Then, in sentence 13, the "star-struck" emoji emphasizes the excitement already expressed in the message by the verbal content and by the capital letters and the elongation in the word "SOOOO". Emoji used for emphasis often have another secondary function (Marcoccia & Gauducheau, 2007; Skovholt et al., 2014), such as politeness in example 12 and the expressive function in example 13.

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

(13) *SOOOO EXCITED FOR THE CONCERT* 🤩

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

Then, emphatic emoji can also strengthen the value of other identical emoji in a message (Cramer et al., 2016), like in example 14. The 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). In their paper on emoji as "beat gestures", McCulloch and Gawne (2018) stated 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, putting your thumb up and 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 appear to work when linked to another unit which carries the emphasized value, whether this unit is a word or another emoji (Schneebeli, 2017).

1.3.6. The structural function

As mentioned in chapter 2, emoji sometimes play the same role as punctuation marks. Early CMD researchers compared their structural function to that of laughter in face-to-face speech, because laughter is able to "punctuate" oral conversation (Provine et al., 2007). Emoji are often found in final position of messages, similarly to full stops ("."), and can be used to indicate the end of a message. To a lesser extent, emoji can also be used between clauses inside a single message (Markman & Oshima, 2007). They are an efficient way to mark a topic shift inside a speaker's conversational turn, as in example 15 below. Sampietro (2016) observes that messages in which emoji occur almost never contain regular punctuation marks (except for question and exclamation marks), suggesting that emoji are good functional substitutes for some punctuation marks in CMC. Emoji can thus be seen as convenient tools for conversational management (Cramer et al., 2016). Like punctuation marks, they can work as syntactic markers inside a sentence or inside a message, signaling different structural boundaries depending on their position (Amaghlobeli, 2012; Spina, 2018).

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

1.3.7. The referential function

The previous functions could almost only be filled by particular categories of emoji. It would be unlikely to find an "elephant" or "bow and arrow" emoji used with the expressive or with the politeness functions, for example. Those functions are usually restricted to facial, gesture and people emoji, or specific "Symbols" emoji (i.e. mostly hearts). However, the referential function can be assumed by virtually any type of emoji.

(16) *Wanna get some* 🌮 *for lunch?*

(17) 👍 *if you enjoyed the video*

(18) *Always in my* ❤️

(19) *I* ❤️ *you*

Referential emoji, also called representational emoji (Beißwenger & Pappert, 2019), are emoji used as referents of the concept that they represent. They can be used as substitutes for a word that they refer to (Amaghlobeli, 2012; Cramer et al., 2016). In example 16, the "taco"

emoji stands for the noun “taco”, whereas in sentence 17 the “thumbs up” emoji replaces the verb “like”. Some emoji are able to represent different concepts, such as the “red heart” emoji that refers to the noun “heart” in example 18 and to the verb “love” in 19. In all these cases, emoji are used as lexemes on the syntagmatic axis and they are part of the propositional content of the message (Schneebeli, 2017).

Such emoji can also be used without verbal content, 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 “emoji lexicon” (McCulloch, 2019). Example 20, an extract of Fred Benenson’s (2010) “emoji translation” of Moby Dick, illustrates how hard it is to understand an emoji story without its literal translation. Benenson conceded that “there are large portions of the book that don’t make any sense” (Goldmark, 2014).

(20) 👨💡👍👨🐋🌀👯👯

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

(Benenson, 2010)

(21) *christmas is almost there!* 🎄🎅🎁

Some referential emoji are also used in messages in which the concept represented by the emoji is already present in the verbal content of the message. In example 21, the three emoji “echo” the referred concept (i.e. Christmas) through an iconic repetition (Na’aman et al., 2017). Moreover, many 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 and would require further investigation.

1.3.8. Extralinguistic functions of emoji

One of the “extralinguistic” ways in which one can use emoji is for an aesthetic purpose. It is probably an important reason why many referential emoji are used, as they make the message more colorful by acting as decorative elements (Cramer et al., 2016). Emoji can also be used in an aesthetic way in order to make a message more attractive. On social media threads

containing many messages or comments from different users, the presence of emoji can make a particular message stand out to the readers' eyes and draw the attention of other users (Yus, 2014; Spina, 2018).

Emoji can also have a ludic purpose (Beißwenger & Pappert, 2019). They can be used in a playful way, for example in sentence 22 below for wordplay: the phonological resemblance between the words "don't" and "doughnut" is exploited. These ludic emoji can also have a relational function, because "perceived playfulness helps to decrease the psychological distance between people" (Tang & Hew, 2019).

(22) *i* 🍩 *like you*

(Na'aman et al., 2017)

(23) 🙂 *do you also think this emoji is creepy?*

Finally, emoji can be used to refer to the emoji itself, as a mention. This simply occurs when someone wants to cite a specific emoji, for example in order to comment its appearance, its function or the people that use it. Example 23 illustrates this.

1.3.9. Emoji position

An important feature to consider when investigating the functions of emoji is their placement inside messages. Few papers on the subject have taken emoji position into account. This criterion could be analyzed by evaluating the distance of an emoji in a message on a scale from 0 (first character) to 1 (final character of the message). Novak et al. (2015) found that the great majority of emoji appear near the end of the message, with an average distance of 0,75. On the other hand, Robert Provine, along with other searchers, identified five potential placements for emoticons in a message: "naked" (emoji alone in a message without verbal content), "before a statement", "after a statement", "at a phrase break in midstatement", and "midphrase in midstatement" (Provine et al., 2007). They discovered that most emoticons appeared after statements or at breaks (supporting the structural function of emoji).

This framework recalls a methodology used in a previous paper, focusing on the analysis of *haha* as a discourse marker in French and English CMD (Escouflaire, 2018b), inspired by Uygur-Distexhe's (2012) research on the placement of *lol*. The seven positions identified for *haha* in CMC messages were the following: "naked", "initial", "final", "medial between two clauses", "medial inside a clause", "quasi-initial", and "quasi-final". Those positions revealed

to be correlated with some of the functions of *haha* as a discourse marker. Though this typology resembles the one provided by Provine et al. (2007), it seems to be more complete and more accurate. Therefore, it could easily be applied to emoji position.

1.3.10. Summary

The different functions that emoji can have in CMC were outlined in this section. Some of these functions, however, are more common than others, and some appear more often in particular genres of CMD. Cramer et al. (2016) investigated the use of emoji in private conversations and found that the expressive function was by far the most popular function for emoji, followed by the interpretative and the relational function. Regarding social media, an analysis of Facebook comment threads by Herring and Dainas (2017) showed that expressive and interpretative emoji were the most frequent functions, then followed by referential emoji.

Functional analysis of emoji is a new field of research where much has yet to be discovered. The new typology that was presented in this chapter aims to help unveiling the different ways in which emoji are used by all types of CMD users, be they male or female, old or young.

1.4. Sociological variation in emoji use

Sociolinguists have studied a great variety of the aspects of CMD. Similarly to the fact that, depending on several factors, people do not speak in the same way in face-to-face conversation, the same applies to written online conversation (Androutsopoulos, 2006). Aside from the technological factors that can influence the way people use language (such as the particular device or website on which the users are communicating), situational factors are also important: the topic of the conversation, its tone (e.g. serious or playful) and the relationship between the interlocutors are some of these contextual variables that are important to take into account when analyzing CMD. (Herring, 2007). However, some of these social factors in particular, which are related to the specific features and attributes of the users, have been proven to have an interesting impact on the production and the interpretation of emoji use. These factors include the gender and the age of CMD users (Dresner & Herring, 2010).

1.4.1. Emoji and gender

Several studies have investigated the influence of gender on emoji production. Regarding emoticons, the "primitive" form of emoji, Alecia Wolf (2000) discovered that they tend to be used more by women than by men, which was confirmed by other research projects on emoticon variation (Tossell et al., 2012; Nishimura, 2015). Wolf (2000) also noticed that female users are more likely to use emoticons with an expressive purpose (for displaying emotions), whereas men use them more often with the interpretative function. Even though Wolf admits that these findings seem to support the stereotype of the passionate woman and the pragmatic man, it appears that her results are strongly dependent on the context of production. In fact, in group discussions in which there are more men than women, female users tend to use fewer emoticons, while men use them more in contexts where women are dominant. Interestingly, in mixed groups, men and women seem to produce emoticons with roughly the same frequency.

Another investigation was conducted on 1200 French private text messages in Belgium (Escouflaire, 2018c). The findings confirm that slightly more emoticons are produced by female than by male users. Some emoticons appear to be more frequently used by men (e.g. « :) »), others more by women (e.g. « :P »). An interesting observation is that men are more likely to use the "hyphenless" forms of emoticons than women (« :) » instead of « :-) »).

Then, a study was carried out in 2017 on the impact of gender on emoji production (and not emoticons). The results, based on 400 million private messages, indicate that emoji, like emoticons, are used more frequently by women than by men (Chen et al., 2017). It also seems that the most commonly used types of emoji are not the same for male and female users. Specifically, while the most popular emoji seem to be produced by both men and women, some particular emoji are more popular among a specific gender. For example, the "party popper" emoji is a lot more likely to be used by women, and the "cigarette" emoji is more frequent with men. An interesting finding of this study is that facial emoji (from the "Smileys & People" category) are usually more popular among female users, whereas emoji representing hearts are more frequently produced by male users. The authors of the study suggest that this difference is due to the fact that men are less likely to express love in real life, thus they use CMD and emoji to communicate their feelings more often than women. These results have also proven to be efficient for inferring gender: a machine learning algorithm was developed to successfully predict the gender of users based on their production of emoji, with the same accuracy as similar text-based algorithms (Chen et al., 2017). However, some researchers observed in another

study, based on Chinese text messages, that gender did not appear to have any impact on emoji production (An et al., 2018). The influence of gender on emoji use might thus be dependent on other factors, such as the user's culture or the language, which will be investigated in section 1.5.

Gender influence has been studied not only in relation to emoji production, but also regarding emoji interpretation. For a study on the linguistic representations conveyed by emoji, 576 participants were asked to tell whether particular examples of text messages containing some emoji, a lot of emoji or no emoji at all, were written by men or women (Escouflaire, 2018a). According to the results, the presence of emoji in an anonymous message is considered by many as an indicator that the user is female. Additionally, the more emoji a message contains, the more people think that it was produced by a woman. Baron and Ling (2011) interviewed American adolescents on how different they thought boys and girls wrote text messages. It appeared that female users were considered as the main producers of "smiley faces", whereas boys were considered less expressive when texting with their female peers.

Another question on emoji interpretation which should be considered is the following: if men and women use emoji in different ways, do they also understand them differently? A recent survey conducted on 628 people found that there is no significant difference in emoji interpretation between male and female users (Herring & Dainas, 2018). The only small divergence occurred on specific emoji functions. For instance, the politeness function is more frequently associated with the "slightly smiling face" emoji by men, whereas women are more likely to link the same function to the "winking face" emoji. Interestingly, the people who did not identify themselves as either male or female showed very different interpretations of many emoji. Herring and Dainas (2018) suggest that this could be linked to their overall younger age compared to the majority of the male and female respondents. The influence of age on emoji production and interpretation is discussed in the next section.

1.4.2. Emoji and age

Before addressing the influence of age on emoji production, some researchers have investigated the different ways in which "younger" and "older" people used emoticons. Nishimura (2015) observed that Japanese CMD users under the age of 40 use a lot more emoticons than the others (more than 40 years old). However, gender still seems to be a more influential factor than age (in Japan, at least), as senior female users tend to produce more emoticons than male users under 40. In French text messages from Belgium, the users who are

the most likely to use emoticons are those between 16 and 25 years old (Escouflaire, 2018c). After the age of 30, both men and women appear to use emoticons a lot less frequently. Yet, it should be pointed out that these results were based on text messages sent in 2004, more than fifteen years before this thesis is written, meaning that these observations may be outdated. Moreover, emoji have nowadays become way more popular than emoticons.

A study on the different factors influencing emoji production found that people above the age of 50 produce significantly fewer emoji than younger users (An et al., 2018). Users less than 35 years old also produce a wider variety of emoji types, while also using them with more various functions than the older users. An interesting example is the case of the "slightly smiling face" emoji, "🙂", which is always used to convey positive emotions by people above the age of 25, but can be used in more complex ways by younger users, in order to display irony or even with a negative emotional value. In another study, focusing on "how older people use emoji" (Gallud et al., 2018), the findings are quite different. The survey, answered by 95 participants, showed that age is not an influential factor for emoji production. People of all age groups stated that they use emoji on a regular basis. However, it was also observed that users under the age of 25 produce more different types of emoji and more complex emoji than their older counterparts.

A factor that has to be taken into account when analyzing emoji production in relation to age, is the experience of the users with emoji and with CMD in general. As explained by Alshenqeeti (2016), the more someone is exposed to messages containing emoji, the more likely he or she is to understand their meaning and to use them in turn. Young people, especially those under the age of 25, have grown up along with the rise of social media and CMD, and have thus naturally become the most productive users of emoji.

Regarding the linguistic representations that emoji convey on their user's age, the survey conducted in Belgium on French-speaking users yielded interesting results (Escouflaire, 2018a). When looking at a message containing emoji, most participants thought that the writer of the message was between 15 and 30 years old. This tends to support the results of other studies on emoji production which determined that young people, and more specifically young female users, are the most productive users of emoji (Nishimura, 2015; Chen et al., 2017; An et al., 2018). However, some younger respondents, when asked to review a message containing more than three emoji of the same kind, answered that they imagined the producer of said message to be over the age of 40, who "tried to look young". The abundance of emoji in a message was thus considered by some young proficient CMD users as a failed attempt to overcome the lack of experience with emoji. This was already acknowledged by Wolf (2000),

who observed that the overuse of emoticons could be a blatant mark of inexperience. Another feature associated with older or inexperienced users was the use of emoji alongside simple punctuation, which some participants considered redundant (Escouflaire, 2018a). This indicates that the structural function of emoji might be seen as a dominant one by younger users of CMD.

1.4.3. Other factors influencing emoji use

As seen throughout this section, the influence of both the age and the gender of users on emoji production and interpretation have been investigated by many different researchers. Overall, emoji appear to be more frequently associated with female users, and women are slightly more likely to use them than men. Age has a less significant influence than gender on emoji use, but some studies have shown that the younger samples of given populations are usually more productive users of emoji.

However, it should be taken into consideration that all the studies on emoji production discussed in this section analyzed messages produced in a particular communicational context: private messaging. The use of emoji in public contexts, such as on social networking sites, has not yet been explored. CMD on SNSs involves that messages can be read by many other people, and the relationship between the writer and the reader of a message can be an important factor in linguistic variation (Androutsopoulos, 2006; An et al., 2018).

In addition to this, other social factors than age and gender have also been identified as having an impact on the use of emoji. The relationship between the interlocutors (An et al., 2018) and the user's experience with CMD and with emoji (Gallud et al., 2018; Escouflaire, 2018a) can have an important influence on the way they use emoji, as mentioned above. However, the cultural background of the users and the language in which the messages are written are undoubtedly essential matters to tackle whenever emoji variation is discussed (Alshenqeeti, 2016). These issues will be examined in the following section.

1.5. Emoji across culture and language

As seen in section 1.4., features specific to individual users, such as age and gender, can influence their use of emoji. However, some traits of the community in which an individual is included can also have an impact on that user's production and interpretation of emoji. These elements are related to the different cultures inside of which users of CMD evolve and live.

Outside CMD, in face-to-face conversation, people from different cultures do not use facial expressions to display emotion in the same ways (Jack et al., 2012). Consequently, considering that most emoji, and before then all emoticons, are iconic representations of facial expressions, it has been suggested that they can be used or interpreted with different meanings in different cultures. First, several studies looked into the differences between emoticons and kaomoji. Prior to the emergence of emoji, emoticons such as ":-)" were really popular in Western countries (Europe and America, mostly), while kaomoji like "^_^" were almost exclusively used in the East (primarily in Eastern Asia). Park et al. (2014) state that this difference is linked to the fact that Westerners interpret facial emotions focusing on the muscles around the mouth, while Easterners are used to read them from the speaker's eyes. Similarly, in CMD, emoticons differ mostly depending on the "mouth" part (e.g. ":-(", ":-P" or ":-D"), whereas the "eyes" part is usually the varying elements for kaomoji (e.g. "o_O", "T_T" or "X_X"). After further examination, the study suggests that the use of emoticons or kaomoji is not influenced by geography, but by a particular dimension of a country's culture: individualistic cultures tend to use the "mouth-based" emoticons, and collectivistic cultures seem to prefer the "eye-focused" kaomoji (Park et al., 2014). Another study highlighted that language is also more influential than geography in emoji production (Park et al., 2013). For example, users living in Eastern countries used kaomoji in Tweets written in their native language, but used emoticons when switching to English. This indicates that the production of emoticons or kaomoji might be strongly dependent on the language used. Moreover, Markman and Oshima (2007) found that emoticons and kaomoji did not share the exact same range of linguistic functions, with kaomoji displaying more complex and varied uses than emoticons.

However, since 2014, emoji have become more popular than both emoticons and kaomoji all over the world (Pavalanathan & Eisenstein, 2015; Ljubešić & Fišer, 2016). They are used in CMD in many different countries and languages, which has led some researchers to refer to them as a "ubiquitous language" (Lu et al., 2016). Chen et al. (2017) developed an emoji-based model which proved to be successful in determining the gender of a user depending on the emoji used in his or her Tweets. Unlike similar text-based models, which are usually

built for one particular language, this emoji-based algorithm worked for messages in almost all of the languages it was applied to. Yet when tested on Arabic Tweets, the algorithm was not as efficient. The authors attribute this difficulty to cultural reasons: the Arabic society's "stricter self-censored emotional behaviors could have on effect on self-presentation online". This suggests that emoji production is significantly influenced by culture and language.

Lu et al. (2016) examined 427 million instant messages containing emoji, sent from 212 different countries and regions of the world. 1281 types of emoji were found in the corpus, with 119 of these representing 90% of all occurrences. Emoji are thus spread following a typical "power-law" distribution, with a small portion of varieties of emoji being very frequently used throughout the corpus, while the majority of emoji types only appear a few times. Looking at differences across countries, one interesting observation is that users in France produce a lot more emoji than those from other regions. Moreover, while facial emoji are the most popular emoji in most parts of the world, French users appear to be more likely to use emoji representing hearts than faces. Then, countries were clustered in groups depending on their emoji preferences. The results show that most European countries share similarities in emoji production, just like South American countries which are also grouped together. This supports the claim that regional and cultural factors have an influence on emoji production. Interestingly, Spanish users were placed in the South American cluster, suggesting that language also has a significant importance. Finally, the authors state that the fact that Russian and Indian users were paired together confirms this linguistic hypothesis, as both Russian and Sanskrit historically belong to the *satem* Indo-European linguistic group. The final results of the study also match with the conclusions of Park et al. (2014) on emoticons and kaomoji: in individualistic societies, CMD users seem more likely to use emoji than in collectivistic societies (Lu et al., 2016).

(1) It's flipping hot out here! 🔥

(2) Inciamos el nuevo año con illusion! 🔥

(Let's start the new year with enthusiasm! 🔥)

(Barbieri et al., 2016b)

Barbieri et al. (2016a) narrowed the focus down to emoji in 30 million Tweets written in only four languages or dialects: Italian, Peninsular Spanish, American English and British English. They compared the production and meaning of emoji across those four languages and made some relevant observations. First, the most popular emoji are very similar in all

languages. However, when looking at particular emoji which are less frequently used, it seems that they behave in an idiosyncratic way in some languages. One specific emoji can be used with a very different function or meaning depending on the language in which it is used (Barbieri et al., 2016b). A good example of this is the "fire" emoji. In English Tweets of the 2015 corpus, as in sentence 24 above, it mostly seemed to act with a referential function (with the fire referring to the word "hot"). On the other hand, in Spanish Tweets, as in example 25, the "fire" emoji often had an expressive or emphatic function, conveying sheer excitement.

Other examples of these cross-linguistic differences in emoji usage are found by Barbieri et al. (2016a), when investigating the most common emoji pairs in each language. In American English, the "waving hand" emoji, "👋", is often combined with hearts or facial emoji (the "face blowing a kiss" or "winking face"): it thus carries the same meaning as farewell expressions such as "goodbye". In British English, the same emoji is found alongside emoji representing flags, trains and landscapes, linking it to the lexical field of travel. Cultural influence seems to have importance, particularly with referential emoji: for example, the "pizza" emoji ("🍕") is often associated with the "fork and knife" emoji ("🍴") in Italian, but very rarely in American English Tweets. Similarly, the "water wave" emoji ("🌊") frequently appears alongside the "sun" emoji ("☀️") in Italian and Spanish Tweets, but not in British English messages. However, the authors conclude by admitting that, even though some emoji behave differently from one language to another, all the most frequent emoji seem to keep their popularity and their meaning across the four languages.

Then, a global analysis of emoji by Ljubešić and Fišer (2016) showed that emoji in Tweets were a lot more popular in South America and South-East Asia than in other regions of the world. Surprisingly, Japan and the United States of America had some of the lowest rates of Tweets containing emoji. They also found that "first world" countries (i.e. economically stable countries) were less likely than others to use facial emoji, and used referential emoji more frequently. Novak et al. (2015) focused on emoji representing clear positive or negative emotions (mostly facial emoji and hearts) in Tweets and observed no significant differences in popularity between 13 different languages.

In the end, it appears that both the culture and the language of CMD users have at least some influence on their use of emoji and on how they interpret them. However, an important question is: to what extent are culture and language significant factors of influence in emoji production and interpretation? It will be very interesting to add a cross-linguistic dimension to

the analyses of emoji usage in this thesis. At the same time, social factors like the age and gender of users (as seen in the previous chapter) can also influence the way people use emoji. Another matter to consider is the temporal aspect of emoji study (Park et al., 2013; Lu et al., 2016). Are they used today in the same way as they were used a few years ago? Most studies presented in this chapter analyzed data from the years 2014 to 2016, it is thus likely that emoji usage has significantly evolved since, as CMD evolves at a very fast pace. A final limitation related to these research works is that they mostly investigated emoji in Tweets, and users of Twitter (in the United States, at least) are usually more educated and wealthier than the general population (Smith, 2011; Park et al., 2013). Emoji usage has not often been compared across several social networking site (SNS) or CMD platforms, it would thus be relevant to investigate if they are used in different ways from one platform to another (Lu et al., 2016).

In the next chapter, the methodology used in this thesis will be outlined. After an overview of the research questions and hypotheses, the data used for the analysis and its collection will be detailed, as well as the methods applied for examining the corpus.

2. Methodology

2.1. Research questions and hypotheses

As explained in detail in chapter 1, emoji have been studied by different linguists interested in several aspects: the way they evolved, their linguistic functions, or how they vary depending on sociological and linguistic factors. Many of these studies inspired the research questions that will be investigated in this thesis.

Even though multiple typologies of the linguistic functions of emoji exist, no consensus has yet been reached among researchers. The roles of emoji in computer-mediated discourse are complex. Some typologies only distinguish between two or three main functions (Cramer et al., 2016; Na'aman et al., 2017), while others identify eight or more different roles for emoji (Beißwenger & Pappert, 2019). With the methodology presented in section 1.3., we decided to build a framework composed of seven linguistic functions: expressive, interpretative, relational, politeness, emphatic, structural and referential. This typology was based on previous studies as well as on personal observations. However, it needs to be tested and verified in order to make sure that it is operational for categorizing emoji on the basis of their functions.

For this reason, the primary objective of this thesis will be to test our "7-functions typology" on a large corpus of emoji used in context. Applying this framework on such a dataset can have several potential outcomes. First, it could be highly successful and demonstrate that the typology is totally operational. The typology could also prove to be dysfunctional, 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, requiring some modifications in order to become operational.

In the previous chapter, it was shown that emoji production can vary depending on different factors. The gender, age and cultural background of the user, the CMD platform used, the topic of discussion and the language of conversation are some of the most important – and most studied – aspects of emoji variation. The "emoji corpus" on which the 7-functions typology will be tested will have to account for these different variational dimensions. This way, it will be verified whether our typology is operational for emoji on different platforms, on different topics, in different languages and for users with different sociological and cultural backgrounds.

The functional examination of this large corpus of emoji in context will also be useful for comparative analysis of emoji production across various socio-technical factors. The results

of this analysis could subsequently support or contradict results on similar subjects from previous studies mentioned in the state of the art.

The two main research questions which will be investigated in the next chapters of this thesis are the following:

- RQ1. Is the "7-functions typology" for the analysis of emoji in context operational? If not, how can it be improved?
- RQ2. Do the linguistic functions of emoji in context differ depending on:
    a) the CMD platform?
    b) the topic of discussion?
    c) the age and gender of the user?
    d) the language of production?

  And if yes, how do they differ?

Before answering these two questions, some expectations can be suggested regarding the upcoming results. These are presented in the form of two hypotheses:

- H1. The "7-functions typology" needs improvement: some functions will be proven unnecessary after the corpus analysis, and new important functions will emerge.
- H2. The linguistic functions of emoji will differ significantly depending on the CMD platform, the topic of discussion, the age and gender of the user and the language of production.

In the next section of this chapter, the corpus used for answering the two research questions will be detailed. After explaining how the data was selected and where it was collected from, the methods used for the analysis of the corpus will be outlined.

2.2. Data

2.2.1. Corpus compilation

a. A multimodal corpus

The corpus used for testing the 7-functions typology has to fulfill a number of criteria. First, the data has to consist of emoji in context, i.e. emoji used alongside other units of language. Therefore, the corpus will be multimodal, as it will consist of both emoji data and text. Indeed, the textual context is often required in order to understand an emoji's function. For this reason, particular kinds of emoji were deliberately excluded from the material:

- "Naked" emoji. They are emoji used in a message on their own, in a standalone way. These emoji do not appear alongside text or with other emoji, which can make it very difficult to interpret them correctly.
- Stickers (only on Facebook). Like emoji, they are "graphicons" (Herring & Dainas, 2017), and some of them look very similar to emoji. However, they are quite bigger and do not fit in the Unicode Standard, unlike emoji. Most of the time, stickers are used on their own in order to react to a message or a previous comment.
- "Facebook Reaction" emoji. In 2016, Facebook gave its users five new "reaction" buttons, in addition to their "Like" button: "Love", "Haha", "Wow", "Sad" and "Angry". In April 2020, during the coronavirus crisis, the "Care" reaction was also added by Facebook.[11] Although these buttons take the shape of existing emoji, they can only be used by users to react to a post or comment and cannot be accompanied by text or by other emoji (Larsson, 2017).

The best way to collect emoji in context for our multimodal corpus is thus to find emoji used alongside text in messages, posts or comments.

b. Data source

The second criterion is the accessibility of the data. Emoji used in instant messaging conversations or text messages are very hard to collect, as most instant messaging applications (Facebook Messenger, Viber…) are used for private conversations, and text messages are also private. Public CMD data can be found in large quantity in blog articles, on forums or on social

[11] https://www.theverge.com/2020/4/17/21224805/facebook-care-reaction-like-button-messenger-app, last accessed on May 30, 2020.

networking sites (SNSs). However, forums often use their own "lexicon" of non-Unicode emoji, and most blog writers do not use emoji at all. Therefore, the choice was made to collect data from SNS, as they consist of public data (partially, at least) and because emoji are used quite often by their users.

Then, research question 2 requires the data to come from two or more different CMD platforms in order to allow for comparison and to test whether the 7-functions typology is similarly operational across different platforms. For this reason, it was necessary to collect data from at least two different SNSs. After examining the ten most popular SNSs in the world (Ortiz-Ospina, 2019), it was decided to use data from Facebook and Twitter. The other popular SNSs were excluded for various reasons:

- WhatsApp and Snapchat are primarily used for private conversations or group discussions; their data is not public.
- WeChat and Weibo are almost only used in China, and their users do not usually communicate in English nor in French, which are the two languages that will be investigated in this study (as explained further in this section).
- Reddit and Tumblr users almost never use emoji in their messages. On Reddit, the use of emoji is not seen as an acceptable linguistic norm by many of the website's users. An entire subreddit called "Emoji Police" is even dedicated to stigmatizing the use of emoji on Reddit.[12]
- Instagram and Tiktok are SNSs on which emoji are really popular (which is likely linked to the overall young age of both apps' users). However, it is very hard to collect data from these two websites, as they revolve mainly around pictures and videos. Moreover, a lot of the emoji used on Instagram and Tiktok are "naked" emoji, used simply to react to images and videos.
- Similarly, Youtube (or at least the comment sections of the website's videos) is a platform on which the use of emoji is quite frequent, but they seem to be mostly used in order to react to videos. Therefore, the emoji on YouTube are less frequently used for conversational purposes than those produced on Facebook and Twitter.

Twitter was selected as the first SNS in this study because it allows for fast extraction of data. The "advanced research" function of the website allows for the easy collection of messages about a specific subject, from selected users, or written in a particular language, region of the world or time span. On Twitter, people debate and have conversations about very

[12] https://www.reddit.com/r/EmojiPolice/ (last accessed on May 30, 2020).

diversified subjects (e.g. the news, culture, everyday life). This suggests that emoji can be used with a wide range of different functions on the website. Then, all of the messages posted on Twitter (called "Tweets") are public, which means that they can technically be used for such a study without the speakers' authorization. Of course, the names of the users whose Tweets are analyzed will be anonymized. Twitter was also selected because it has been used in many previous studies on CMD and emoji, allowing for the efficient comparison between our results and those of earlier studies.

Facebook, on the other hand, was chosen for multiple reasons. With 2.32 billion monthly users at the start of 2019 (Wilberdin & Wells, 2019), Facebook is by far the most popular SNS in the world. The website is used every day by people all over the world. Users go on Facebook to express themselves and to comment on many subjects. "Facebook pages" and "Facebook groups" allow for the collection of data centered on particular subjects or from specific areas and groups of people. Finally, even though a lot of the messages shared by its users are private, there are some exceptions. This is where the messages we analyze will be collected from:

- Posts and comments written on public Facebook pages and public Facebook groups, which can be accessed by any Facebook user.
- Posts and comments written on private Facebook groups containing thousands of users. The messages used in this study which are extracted from such groups will naturally be anonymized, like all other messages.

In section 2.2.2., we will look at the differences and similarities between Twitter and Facebook on the basis of Susan Herring's CMD genre classification scheme (Herring, 2007).

c. Comparable data

The third essential criterion for building the corpus is that it has to be composed of multiple parallel subcorpora in order to allow us to answer the second research question. The data will not only be compared between Facebook and Twitter (different CMD platforms), but also between languages, between topics of discussion and depending on age and gender. Each of the subcorpora should contain the same amount of data, in order to allow for comparison across each of the variables. Here is how the corpus is divided following every factor:

- Language. It was decided to focus on two languages, English and French. These choices were obvious to me, as they are the two languages that I know best and have studied the most. This linguistic proficiency is necessary for analyzing the functions of emoji in context efficiently. English and French are among the five most spoken

languages in the world as of today, non-native speakers included (Simons & Fennig, 2019). For this reason, it is not difficult to find messages written in those languages on Facebook and Twitter. Moreover, English and French are quite similar in many aspects: lexicon, linguistic structure, semantics, etc. These shared points allow for easier comparison between the two languages. British English and Belgian French were chosen as the two dialects which will be investigated in this thesis (when possible) in order to avoid too much variation inside each of the two linguistically comparable subcorpora.

- Topic. Three different subjects of discussion were selected, in order to analyze emoji from different contextual situations: entertainment, politics, and conversations about everyday life. The objective is to find emoji used in serious CMD contexts (like discussions about politics) as well as in more trivial situations (such as small talk). The "entertainment" context serves as an intermediate ground. These subjects were also chosen because of how easy it is to find conversations about these topics on Facebook and on Twitter, and in French and in English.
- Gender. The decision was made to focus only on male and female, because it is really difficult to infer other non-binary genders from the information available when collecting messages from SNS users. The number of messages sent by male or female users will probably not be exactly equal, thus the corpus will not be exactly parallel on this factor. However, it will give us a good overview of the differences in emoji use between users of the two genders. The methods used for identifying a user's gender are detailed further in section 2.2.3.
- Age. Facebook and Twitter do not allow for an efficient extraction of their users' ages. While gender can be (relatively) reliably determined (even more on Facebook than on Twitter), it is not possible to identify with certainty the age of most users on neither of both websites. For this reason, right after the beginning of the analysis, it was decided to remove the "age" variable from our comparative research.

With two different CMD platforms, two languages, three topics and two genders, the entire corpus consists of twelve subcorpora. Each subcorpus is composed of 100 emoji in context, with the messages in which they are used, for a total of 1200 emoji in context. A detailed description of each of the twelve parallel subcorpora is given in the next section.

2.2.2. Corpus description

a. Facebook and Twitter

In 2007, renowned CMD linguist Susan Herring presented a classification scheme for CMD genres. It consisted of a list of different features that can be used for comparing CMD platforms (Herring, 2007). In this case, Facebook and Twitter will be approached using Herring's framework, which was explained in detail in the first chapter of this thesis. However, the researcher revised her previous model a few years later, after the emergence of new modes of communication in the "Web 2.0", as she calls it (Herring, 2013). Herring decided to take into account the phenomenon of "media convergence" into her classification scheme. Media convergence refers to the fact that CMD users are now able to use different modes of communication, different media, at the same time, on one single CMD platform. For example, most instant messaging apps (e.g. Facebook Messenger and WhatsApp) allow their users to send messages composed of text, but also images, videos, GIFs and even voice messages. Emoji also play a role in this new multimodality of the Web 2.0. In a more recent article, Herring (2019) discusses this once again, and concludes that her 2007 CMD classification scheme is still relevant. Technological and social factors are still important nowadays, but the new modes of communication and the increasing media convergence need to be taken into account.

Therefore, Facebook and Twitter are classified in Table 1 below following Herring's 2007 model, but this framework has to be used with caution and in the light of the new paradigms brought by the recent modes of communication in CMD. We will only look at the medium factors (or technological factors), which were all detailed in section 1.1.3. The social factors are almost always dependent on the specific context of the conversation, which is why they cannot be reliably used for comparing two SNSs.

The model for medium factors of CMD genres shows that Facebook and Twitter are not very different on most aspects. Just like many other SNSs today, they are multimodal, communication between users is asynchronous, and the messages appear in a temporal order (a Facebook comment posted at 12h08 will be displayed under a comment posted at 12h05). Yet, there are two major differences which are made clear by the model below. The first one concerns the possibility to communicate anonymously on the SNS. Facebook has a "real name" policy: users who enter the website must declare that the name they use on the website is their legal name, or at least the name they go by in real life. Therefore, most people use their real name when posting on Facebook. Twitter, on the other hand, does not have same policy. Users are free to post under a pseudonym, and therefore many people do not use their real name on

when posting on Twitter. However, most celebrities and politicians (people for whom social media presence is part of the job) use their public name on the website. This difference has many implications on the behaviors of the users.

Table 1. Facebook and Twitter medium factors (Herring, 2007).

Medium factors	**Facebook**	**Twitter**
M1. Synchronicity	Asynchronous.	Asynchronous.
M2. Message Transmission	Messages are received one after the other.	Messages are received one after the other.
M3. Persistence of transcript	Messages can be erased or edited by their sender.	Messages can be erased by their sender, but not edited.
M4. Size of message buffer	63,206 characters.	280 characters.
M5. Channels of communication	Multimodal.	Multimodal.
M6. Anonymous messaging	Most people use their real name.	Many people do not use their real name (pseudonyms).
M7. Private messaging	Allowed, but some Facebook posts and comments are public.	Allowed, but all Tweets are public.
M8. Filtering	"Blocking" feature.	"Blocking" feature.
M9. Quoting	"Tagging" and "hashtag" features.	"Tagging" and "hashtag" features.
M10. Message format	Messages appear in a temporal order.	Messages appear in a temporal order.

Then, the other technological difference is the message size limit. On Facebook, a single post can contain up to 63,206 characters, a very high number which is not likely to be reached by one message in a conversation. By contrast, Twitter restricts its Tweets to a limit of 280 characters, which can be quite short. Before 2017, the Tweet size limit was of 140 characters. This sometimes forces users to express what they want to say in a more condensed and more straightforward way than on Facebook. This difference should thus be taken into account when comparing our results, because the use of emoji, which are one single character each, can be seen as an efficient tool to express ideas in a very short format.

Although the social factors of Herring's (2007) model are not relevant for such a broad comparison (as it would have to be applied to specific contexts of conversation), there are some important social differences between Twitter and Facebook. First, the people who use Twitter are not the same as those who use Facebook. In other words, the population represented by Twitter users is not the same population as that of Facebook users. There are sociological differences between both websites' users. For example, previous research in the United Kingdom has shown that the average Facebook user is younger than the average Twitter user, and that Facebook users have a higher average level of education (Mellon & Prosser, 2017). Studies set in the United States found that Twitter users are more likely to live in urban areas (Mislove et al., 2011) and are wealthier on average than Facebook users (Malik et al., 2015).

Another social difference between the two SNSs concerns the relationships between users. On Twitter, people are very likely to interact with users that they do not know in real life. On the contrary, Facebook users tend to communicate mostly with people from their "real life" networks: family, friends, work colleagues, etc. (Buccafurri et al., 2015) This can have multiple implications on the way users express themselves online. For example, Facebook users may be more cautious with what they say on the website in order to avoid negative repercussions in real life. In contrast, Twitter users may use their anonymity to be more vehement and write things that they would not be able to say in real life without repercussions. However, on many Facebook pages and in large Facebook groups, great numbers of users who do not know each other interact every day, creating a conversational environment more similar to that of Twitter.

All of these differences need to be kept in mind when analyzing and comparing results on Twitter and on Facebook.

b. The twelve subcorpora

The 1200 emoji were distributed across twelve different subcorpora, with 100 emoji per subcorpus. All of them were found inside messages written between January 6, 2020, and March 30, 2020. Naturally, all of the messages collected contained at least one emoji. In Table 2, metadata on the twelve subcorpora is provided, ranked in the temporal order in which the messages were collected. The content of each subcorpus is briefly detailed below.

Table 2. The 12 subcorpora.

N°	Subcorpus ID	Language	Platform	Topic	Production period	Collection period
1	CFE	English	Facebook	Everyday life in a British city	6/01 → 13/01	10/01 → 13/01
2	CFF	French	Facebook	Everyday life in a Belgian city	9/01 → 16/01	15/01/20 → 16/01/20
3	NTE	English	Twitter	Netflix	10/01 → 17/01	17/1
4	NTF	French	Twitter	Netflix	20/01 → 22/01	21/01 → 22/01
5	NFE	English	Facebook	Netflix	01/02 → 08/02	08/02 → 10/02
6	NFF	French	Facebook	Netflix	22/01 → 24/01	08/02 → 10/02
7	PTE	English	Twitter	Brexit	30/01 → 10/02	11/02 → 12/02
8	PTF	French	Twitter	Brexit	01/01 → 10/02	14/02 → 15/02
9	PFE	English	Facebook	Politics in the UK	20/03 → 24/03	24/03
10	PFF	French	Facebook	Politics in France	21/03 → 26/03	25/03 → 26/03
11	ITE	English	Twitter	Everyday life during coronavirus isolation	21/03 → 28/03	27/03 → 28/03
12	ITF	French	Twitter	Everyday life during coronavirus isolation	28/03 → 29/03	29/03 → 30/03

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 policitians.

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.

2.2.3. Corpus analysis

a. Data 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. First, a total of 11 different variables were identified, constituting 11 columns containing information about the emoji or the message in which it was found. These are listed below, from column A to K.

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 thesis.

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 thesis 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 amount 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* (TRUE or FALSE)

⇨ 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.

(1) n672 & n673: "Now I understand #Brexit 😳😡"

Then, all emoji and the messages in which they were found were analyzed more closely one by one. Three more pieces of information were determined for each emoji after their analysis in context: *gender*, *position* and *function*. These are presented below (from L to N). The different issues raised by the analysis are then detailed in the next section.

L. *gender*

⇨ The gender of the user who posted the message containing the emoji. It can be tagged as *M* (sent by a male user), *F* (sent by a female user) or *unknown* (when the user's gender cannot be reliably determined).

M. *position*

⇨ The position of the emoji inside the message. Five different positions were found to be relevant for classifying emoji, inspired by Uygur-Distexhe's (2012) framework: *initial, between* (for "between two clauses"), *inside* (for "inside a clause"), *final* and *quasi-final* (examples of quasi-final emoji are presented in section 3.1.3). However, some less frequent and problematic positions are discussed further in this section.

N. *function*

⇨ The linguistic function performed by the emoji in the message. As a single emoji can perform more than just one function, up to three different functions were sometimes assigned to one emoji (*function1, function2* and *function3*). The seven different functions are those that were selected for this thesis, as presented in Section 1.3.: *expressive, interpretative, relational, politeness, emphatic, structural* and *referential*. Issues regarding the identification of these functions are discussed in the following section.

b. Annotation issues

Analyzing the 1200 emoji one by one took almost three months of work. During this period, several methodological problems were encountered, which had to be solved in various ways. The most important decisions that were made are explained below.

- It was expected that the identification of some users' gender would prove to be problematic. As explained in section 2.2.1., it was decided to focus only on male and female users, as other genders are very hard to determine from the information publicly available on Facebook and Twitter. There are four different methods to determine a user's gender without access to private data, which are listed below from the most reliable to the least reliable method:
    - Content of the message. Sometimes, users explicitly mention their gender in the message that was collected (e.g. "As a woman, I believe that...") Assuming that

these messages are not false or ironic, this is the most reliable way to know a user's gender. Unfortunately, this only occurred a handful of times in the corpus.

- First name. Of course, looking at the user's name on Facebook and on Twitter seems a straightforward way to identify his or her gender. However, some people do not use their real name, or use a pseudonym, even on Facebook. Moreover, gender switching is quite common on the Web (Androutsopoulos, 2006), and should be included as part of the limitations of this study. Then, some first names are unisex, like Dominique in French or Hayden in English. The next point can often help disambiguating these cases.
- Profile picture. A look at the user's profile can confirm his or her gender. Many users, mostly on Twitter, do not have a picture of their face as their profile picture, so this is not the most reliable method. On Facebook, looking at some of the user's publicly shared pictures can also sometimes help with gender identification.
- Grammar. Sometimes, the user's gender can be inferred through grammatical cues found inside the message. In example 2 below, the French word *obsédée* is used in the past participle form with the suffix "-ée", showing that the speaker is female. This was not found much throughout the corpus, and is not considered as a very reliable method, as spelling errors are quite common in CMD. In addition, this does not work for English data.

(2) n580: "J'ai été traumatisée par l'ep 1 de la saison 4, je voulais tellement la suite que j'ai été complètement obsédée par Outlander une longue semaine durant. 😅

(I was traumatized by the first episode of season 4, I wanted the sequel so much that I was completely obsessed by Outlander during a whole week. 😅)"

If none of these methods could be reliably applied to determine the gender of the user who produced an emoji, then this user's gender was annotated as *unknown*.

- Even though most emoji were found at clear positions in their messages (final, quasi-final, initial, inside a clause or between two clauses), some appeared at more ambiguous positions.
    - First, several emoji were found inside parentheses, for example in message 3 below, in which the "😉" is part of the parenthetical comment. After consideration, these cases were tagged as "between two clauses", as the content in parentheses often works as a single grammatical clause.
    - Then, some messages were clearly divided into different paragraphs (mainly on Twitter). When an emoji appeared at the end of a paragraph, after which another paragraph starts (after a line break), it was also annotated as "between two clauses". The reason for this was that we study the function and the position of the emoji inside the whole message, and not only inside a specific paragraph.
    - A final problem with the identification of emoji position was encountered with emoji appearing at the beginning of a message, right after a tag (i.e. the name of another user mentioned to get his or her attention). As tags and mentions are usually not part of the syntactic structure of the message, they were systematically removed from the raw message (as explained in section 2.2.1). However, in messages such as example 4 below (in which the name of the person tagged has been changed for anonymization), removing the tag puts the emoji in initial position, which is not its original position. Therefore, in these cases, it was decided not to remove the tag, in order to leave the emoji in its original position (in the case of example 4, "between two clauses").

(3) n112: "et de plus, il a déjà "menacé" la ville de devoir prendre ses responsabilités (donner des aides en gros 😉) quand l'argent du mystérieux donateur serait épuisé.... Belle manière de fonctionner !!!"
"(and more, he has already "threatened" the city to take his responsibilities (giving contributions roughly speaking 😉) when the money of the mysterious donor will be used… Nice way to function !!!)"

(4) n446: "**Jane Doe** 😍 12th February woowoooowooooooo"

- Determining the function of an emoji on the basis of our 7-functions typology was very difficult for some cases. During the annotation phase, it rapidly became clear that many emoji have more than one function, and that distinguishing the primary function of an emoji from its secondary (and possibly tertiary) function(s) can be complicated. Some emoji were tagged as having a "blurry function" and had to be analyzed in more depth later. In the end, this led to reconsiderations and to some changes that had to be made to the initial typology. This will be explained in detail in the "discussion" chapter of this thesis (i.e. chapter 4).

  As the identification of an emoji's function(s) can be very delicate, it was necessary to verify whether the choices I made as the only annotator were not too subjective. Therefore, a second person was asked to annotate a random sample of the corpus, and an inter-rater reliability test was computed. This process is explained in the following section.

c. Inter-rater reliability

Using Excel, a sample of 50 emoji, along with their messages, was randomly selected from the corpus. The "7-functions" typology was explained in detail to Emeline Verhelst, an Anthropology Master student proficient in both French and English, who then determined the functions of the 50 emoji on the basis of the typology. She did not have access to any information other than the original messages containing each emoji, in order to avoid skewing her interpretation of the emoji (the platform used and the user's genders were not specified).

It was decided to use Fleiss Kappa in order to compute the inter-rater reliability value, as the annotation of the emoji functions consists of more than two categorical variables (Fleiss & Cohen, 1973). The test was computed using R, in Rstudio (R Core Team, 2019), and gave the following result: $k = 0.61$, $p < 0.001$. With this value, the agreement between the two annotators can be interpreted as "substantial" (Gwet, 2014).

The result of this inter-rater reliability test is relatively high, which means that the functions identified for the 1200 emoji in the corpus can be considered trustworthy. However, there were several divergences between the two annotators on the functions of the 50 emoji in the sample. An interesting difference was found between the two annotations of "💔" in example 5 below. The first annotator determined that the emoji's primary function is expressive, because it is used to show the speaker's emotion, and that its secondary function is referential, because the emoji iconically represents the word "heartbreaking" in the message.

On the other hand, the second annotator identified the expressive function as the emoji's secondary function, and the referential function as its primary function. This problem led to other disagreements between the annotators, which suggests that the difference between primary, secondary and tertiary functions should be more precisely defined. This problem will be further discussed in chapter 4, and a solution will be proposed.

(5) n1096: "love my job as a funeral celebrant. It's such a privilege to help families during their hardest moments in life. But, jeez this week has been tough. Families unable to hug or hold hands and having to sit apart from each other. It's heartbreaking 💔"

It should be noted that the two annotators are around the same age, have the same educational level and are both equally proficient with CMD. It can be suspected that the functions identified would not be exactly similar if the emoji had been analyzed by an annotator with a different sociological profile and background.

d. Methods for statistical analysis

After annotating the entire corpus in Excel, it was possible to analyze the data quantitatively with the help of RStudio and jamovi (The jamovi project, 2020), two software environments facilitating the statistical analysis of large datasets. The first step will be to compute various chi-squared tests for independence in order to investigate how the different factors influence emoji production and the frequencies of the emoji functions. The first variables that will be examined are emoji types and categories, as well as emoji position. Then, the influence of the four external factors mentioned in the second research question (gender, language, topic and platform) will be analyzed using similar statistical tests. Finally, a logistic regression model will be computed to visualize the simultaneous influence of each of these external factors on emoji functions.

Once the statistical results have been described in detail in chapter 3, they will be interpreted and discussed in the final chapter of this thesis, along with many examples from the corpus. The thorough analysis of these examples will add a qualitative layer to our analysis and will provide answers to our research questions.

3. Results

3.1. Preliminary observations

Before digging into the results that will help us answer the two research questions of this thesis, it is important to go through some general observations that were made when analyzing the data. Our multimodal corpus of 1200 emoji in context provided an interesting overview of emoji production on French and English social media. This section aims to explain in detail several aspects of how emoji were used by the 1024 different speakers in the corpus.

3.1.1. Emoji types and tokens

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), this represents barely 6% of the whole emoji "lexicon".

Table 1. 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	🇬🇧	U+1F1EC	16	🧐	U+1F9D0	10
😭	U+1F62D	31	🤦‍♂️	U+1F926	15	🤗	U+1F917	10

Table 1 displays the 30 emoji types that were found the most throughout the corpus, (excluding emoji repetition), and the number of times each of these most popular emoji types was found. It is important to note that 98 of the 195 different emoji types of emoji in the corpus only appear once. These cases of emoji types with a single occurrence in our corpus are 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 1. 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 to emoji studies by several authors who found a similar distribution of emoji types in their corpus (Novak et al., 2015; Lu et al., 2016).

Table 2. The 30 most used emoji types as of March 2019 (Unicode, 2019).

30 most popular emoji according to the Unicode Consortium

Emoji	*Unicode*	*Emoji*	*Unicode*	*Emoji*	*Unicode*
😂	U+1F602	😅	U+1F605	😆	U+1F606
❤️	U+2764	👏	U+1F44F	🙄	U+1F644
😍	U+1F60D	😁	U+1F601	💪	U+1F4AA
🤣	U+1F923	❤️	U+2665	😉	U+1F609
😊	U+1F60A	🔥	U+1F525	☺️	U+263A
🙏	U+1F64F	💔	U+1F494	👌	U+1F44C
💕	U+AF495	💖	U+1F496	🤗	U+1F917
😭	U+1F62D	💙	U+1F499	💜	U+1F49C
😘	U+1F618	😢	U+1F622	😔	U+1F614
👍	U+1F44D	🤔	U+1F914	😎	U+1F60E

It is not surprising that the "face with tears of joy emoji" is the most frequent emoji in our data. As explained in section 1.3., it has been the most popular emoji for years and for several reasons. 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), which is presented in Table 2 above: "❤️", "🤣", "😍", "👍" are really found 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 the corpus were found in public conversations. These two emoji express a high level of intimacy and can be expected to occur a lot more frequently in private conversations. The Unicode list was generated on the basis of

public and private messages. This suggests that the fact that all our messages are public will possibly have an impact on the outcome of the analysis.

In comparison, some emoji, namely "🤔", "😉" and "🙄", are much more frequent in our results than in Unicode's (2019) findings. This could be linked to the political nature of many of the messages in our corpus, some of which involve heated debates. Particular attention should be paid to these three emoji in the investigation of our results. Then the British flag "🇬🇧" is found quite often in discussions about Brexit (i.e. the main topic of two of our twelve subcorpora), which explains its presence in Table 1.

Table 3. The influence of platform, language and gender on emoji type frequency.

	Independent variables						
	Platform		Language		Gender		
Emoji	*Facebook*	*Twitter*	*English*	*French*	*Female*	*Male*	*TOTAL*
😂	57	62	62	57	54	52	119
🤣	32	32	34	30	23	38	64
🤔	31	21	21	31	14	31	52
😍	29	22	33	18	39	8	51
😉	35	13	9	39	10	32	48
❤️	13	27	27	13	24	7	40
👍	22	16	20	18	14	17	38
😁	25	9	19	15	17	15	34
🙄	14	18	20	12	9	18	32
😭	10	21	12	19	16	9	31
	X^2: 328.5***, p = 5.3^{e-09}		X^2: 285.85***, p = 1.9^{e-05}		X^2: 570.6***, p = 4.11^{e-09}		

Table 3 surveys the ten most frequent emoji in the corpus and compares them across three variables (platform, language and gender) A chi-squared test for independence was computed for each variable, and all three were shown to have a significant influence on the types of emoji produced. Some of the emoji presented in Table 3 appear to be more strongly influenced by the three variables. For example, "😍", "❤️" and "😭" are significantly more frequent in English, and are more likely to be produced by female users.[13] They are also significantly more common on Twitter than on Facebook (except for "😍"). On the other hand, "😁" is much more popular

[13] The "gender" column in Table 3 does not include the "unknown" gender emoji, which explains why it has lower totals than the other two columns.

on Facebook. Then, "🤔", "😉" and "🙄" are all significantly more frequent in messages written by males and by French users. Another interesting finding, which is not presented in Table 3, is that the "angry face" emoji, "😠", appears 19 times in the corpus, but only on Facebook. A similar emoji, "☹️", appears 14 times on Facebook and only once on Twitter.

These results indicate that users are influenced by several factors when selecting which type of emoji they will use. This suggests that similar relationships may be found regarding the influence of platform, language and gender on the linguistic functions of emoji.

3.1.2. Emoji categories

There are eight different categories (listed in section 2.2.3.) in which emoji are divided depending on the kind of concept they refer to. Of the 30 most frequent emoji types of our corpus, presented in Table 1, only two emoji are not part of the "Smileys & People" category: "🇬🇧" and "❤️". The frequency of the first one (belonging to the "Flags" category) is due to a recurrent theme of the corpus, as explained in the previous section. The "red heart" emoji, on the other hand, is part of the "Symbols" category. Yet, unlike other "Symbols" emoji found in the corpus, such as "🎶" or "⬇️", hearts emoji are able to convey something that most of the "Smileys & People" emoji can also convey: emotion.

Table 4. The influence of platform, language and gender on emoji category frequency and token frequency.

	Independent variables						
	Platform		Language		Gender		
Emoji category	*Facebook*	*Twitter*	*English*	*French*	*Female*	*Male*	*TOTAL*
Smileys & People 😃	554	474	493	535	458	461	1028
Animals & Nature 🐶	9	19	19	9	14	9	28
Food & Drink 🍎	2	9	9	2	5	4	11
Activity ⚽	1	3	2	2	2	0	4
Travel & Places ✈️	3	2	3	2	0	5	5
Objects 🛠️	3	10	6	7	6	4	13
Symbols ❤️	26	56	54	28	46	16	82
Flags 🇬🇧	3	27	15	15	7	15	30
	X^2: 49.4***, p = 1.9^{e-08}		X^2: 18.2*, p = 0.011		X^2: 50.9***, p = 4.3^{e-06}		

Table 4 above presents the distribution of emoji categories in the corpus and across three binary variables (platform, language and gender). As could be expected, the emoji from the "Smileys & People" category account for most of our data (85%). Next, "Symbols" represent

only 6.8% of all emoji, with most of them being hearts of all sorts: mainly "red hearts", but also "💔", "💙" or "💕", among others. They are followed by the "Flags" (2.5%) and the "Animals & Nature" emoji (2.3%). The unexpectedly high frequency of "Flags" emoji, compared to Unicode's list (2019) in which they are very rare, can be explained by the great number of political discussions found in the corpus. In the "Animals & Nature" category, two emoji appear to be quite popular: "🔥" and "🙈" (both are found five times in the corpus). As explained in section 1.2.3., the "fire" emoji is often used for expressing excitement, rather than for referring to proper flames (Gawne & McCulloch, 2019). The "see-no-evil monkey" emoji represents one of the "Three wise monkeys" in Japanese folklore. It is often used to express embarrassment (Shoeb et al., 2019), as in example 1 below, or with an interpretative function. The four remaining categories ("Objects", "Food & Drink", "Travel & Places" and "Activity") all have a very low frequency in our corpus.

(1) n1059: "58 minute fast bake 🤣, there's butter stains glistening on the floor like slug marks now as I did apply a tad too much 🙈"

Emoji category frequency is highly related to platform and gender, but less significantly influenced by language, as shown by the results of the chi-squared tests in Table 4. "Smileys & People" emoji appear to be more frequent on Facebook than on Twitter, while all the other categories are more commonly found on Twitter (except for "Travel & Places" emoji). This will be investigated in more detail when analyzing the relationship between emoji functions and categories in section 3.2.1.

Then, female users are more likely to use "Symbols" emoji than male users, which is likely due to the higher frequency of hearts produced by women (as shown in Table 3). "Symbols" emoji are also more frequent in English than in French messages.

3.1.3. Emoji position

As seen in Table 5 below, the large majority of the emoji used in the corpus were found at the end of the message (i.e. in final position). This supports the findings of previous studies in which emoji position was taken into account (Novak et al., 2015; Zhao et al., 2018). Then, slightly more than 10% of the emoji in the corpus were found between two clauses. This is

more than what was suggested by the results of Tauch and Kanjo (2016), who found that less than 5% of the emoji on Facebook and Twitter appeared between two sentences.

Table 5. Distribution of emoji positions in the corpus, ranked by total and relative frequency.

	Emoji position					
	Final	Between	Inside	Initial	Quasi-final	Between & final
Abs. Freq.	992	135	28	25	11	7
Freq. %	82.5	11.2	2.3	2.1	0.9	0.6

The fact that almost all emoji are found either in final position or between two clauses can support the claim that emoji can function like punctuation signs (i.e. full stops, commas and semicolons). In example 2 below, the "😂" appears between two clauses, both separating them syntactically and linking them thematically, as a comma would function. This is a good example of an emoji with a structural function. However, some emoji which appear between two clauses are used alongside punctuation signs, as in example 3, in which the emoji is followed by a full stop. A very small number of emoji (0.6%) were also found both in final position and between two clauses, appearing twice in the same message. These emoji, like in example 4, were annotated as "between & final".

(2) n82: "I started there 63 or 64 😂 loved it"

(3) n1043: "All I do is clean, clean, cook, eat and press my phone 😓. This is getting too boring!"

(4) n521: "Entre les Studios Ghibli et Vikings, ça va être chaud début Fevrier 🙂 Top le doc sur Malcom X, une bonne surprise de plus 🙂"

("Between the Ghibli Studio and Vikings, February's going to be hot 🙂 Great, the documentary on Malcolm X, another good surprise 🙂")

Very few emoji were found inside a clause (2.3%) or at the start of a message (2.1%). Only 11 emoji in the entire corpus were "quasi-final" (0.9%). As shown in examples 5 and 6, these 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.

(5) n51: "You was drunk Velda Lancashire doesn't matter if there was another pub 😜 xxx"

(6) n420: "weve not finished the last season! kept falling asleep 😕 xx"

Table 6. Influence of CMD platform on emoji position.

	Emoji position											
	Final		Between		Inside		Initial		Quasi-final		Between & final	
Platform	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
Facebook	516	42.9	55	4.6	7	0.6	6	0.5	11	0.9	4	0.3
Twitter	476	39.6	80	6.6	21	1.7	19	1.6	0	0	3	0.2

$X^2 = 34.14***$, $p = 1.618^{e-05}$

Aside from this, the "language" variable was found to have a very low influence on emoji position. Gender did not seem to affect emoji position either. Table 6, however, shows that the CMD platform on which emoji are produced has a strong influence on their position. On Facebook, emoji are a little more likely to appear at the end of the message than on Twitter. On the other hand, emoji in initial position, between two clauses and inside a clause, are found significantly more on Twitter than on Facebook. This may be linked to differences in emoji functions between the two platforms, as will be outlined in section 3.2.2.

3.1.4. Emoji repetition

Emoji repetition occurs when an emoji appears more than once inside the same message, at the same position and with the same function. Repeating an emoji often strengthens the emotional value or the meaning of that emoji, adding an emphatic aspect to it.

Table 7. Distribution of emoji repetition in the corpus (1 = emoji is not repeated).

	Number of times the same emoji appears							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Absolute Frequency	986	87	93	19	10	2	2	2
Frequency %	82.1	7.2	7.7	1.6	0.8	0.2	0.2	0.2

$W = 0.46$, p-value $< 2.2^{e-16}$

Table 6 shows that emoji repetition did not appear much throughout the corpus. Quite obviously, the data is not normally distributed in terms of emoji repetition, which is confirmed by a Shapiro-Wilk test. Only 17.9% of all emoji were repeated at least once inside their message. When repeated, an emoji usually appears two (7.2%) or three times (7.7%) in the same message, rarely more. A mere 3% of all emoji in the corpus were used four times or more in the same message.

Emoji repetition does not appear to be significantly influenced by any of the variables investigated in this thesis (gender, platform, language or topic). The only interesting finding is

that some emoji types seem to be more likely to be repeated than others. The emoji "👏", "🤣", "❤️" and "👏", in particular, are quite often repeated inside the same message. This will be important for further discussion on emoji functions, more notably on the emphatic function.

3.1.5. Gender distribution

The corpus was divided in twelve subcorpora, so that the 1200 emoji could be separated in groups of comparable sizes, depending on several variables. Therefore, there are 600 emoji from French messages and 600 emoji used in English messages. Likewise, 600 emoji were collected from Facebook and 600 from Twitter. Regarding the topic of conversation, 400 emoji were found in discussions about politics, 400 about entertainment, and 400 in conversations about everyday life.

Table 8. Gender distribution in the corpus and across three variables, platform, language and topic.

	Independent variables							
	Platform		Language		Topic			
Gender	*Facebook*	*Twitter*	*English*	*French*	*Netflix*	*Politics*	*Ev. life*	*TOTAL*
Female	335	203	300	238	228	117	193	538
Male	255	259	240	274	118	224	172	514
Unknown	11	138	61	88	54	59	36	149
	X^2: 140.67***, p = $2.^{2e-16}$		X^2: 14.29***, p = 0.0008		X^2: 74.65***, p = 2.36^{e-15}			

It would have been too complex to collect the exact same numbers of emoji produced by female and by male users. As explained earlier, the user's gender could not be determined reliably in some cases and had to be annotated as "unknown". But even though no particular attention was paid to gender distribution during the data collection phase, it appears in Table 8 above that the emoji in the corpus are distributed quite equally regarding users' gender. Of the 1200 emoji, 538 were produced by women (44.7%) and 514 by men (42.8%). The remaining emoji, for which the user's gender could not be identified, account for 12.4% of the data. For the different reasons outlined in section 2.2.3., identifying a user's gender reliably was much more difficult on Twitter than on Facebook. This is why only 11 emoji from Facebook messages were tagged as "unknown", against 138 emoji in Tweets.

Even though gender is distributed almost equally through the whole corpus, it is clear in Table 8 that it is not the case in the different subcorpora. First, there are many more emoji produced by female users on Facebook, and more emoji from male users on Twitter. Then, the majority of the emoji in English were produced by women. The reverse is the case for French

messages. Finally, regarding topics of discussion, there are more emoji in messages about entertainment (*Netflix* subcorpora) that were written by women, and more emoji in messages about politics that were produced by men.

This unequal distribution of gender in the different subcorpora will have to be kept in mind when analyzing emoji functions across different languages, topics and platforms in section 3.2. There is no doubt that attention should be paid to the influence of this inconsistent gender distribution on the different sections of the corpus.

3.2. The seven linguistic functions of emoji

3.2.1. Distribution of functions in the corpus

a. Functions frequency

The distribution of each of the seven emoji functions in our corpus is presented below, in Table 9. Those seven linguistic functions were explained one by one in detail in section 1.3. Some emoji were identified as having more than one function, which is why the frequencies of secondary and tertiary functions are also listed. However, when analyzing functional distribution in the corpus, the focus should always be put on the primary functions of emoji. The methods used to identify which of the functions of an emoji is the primary function are discussed in chapter 4.

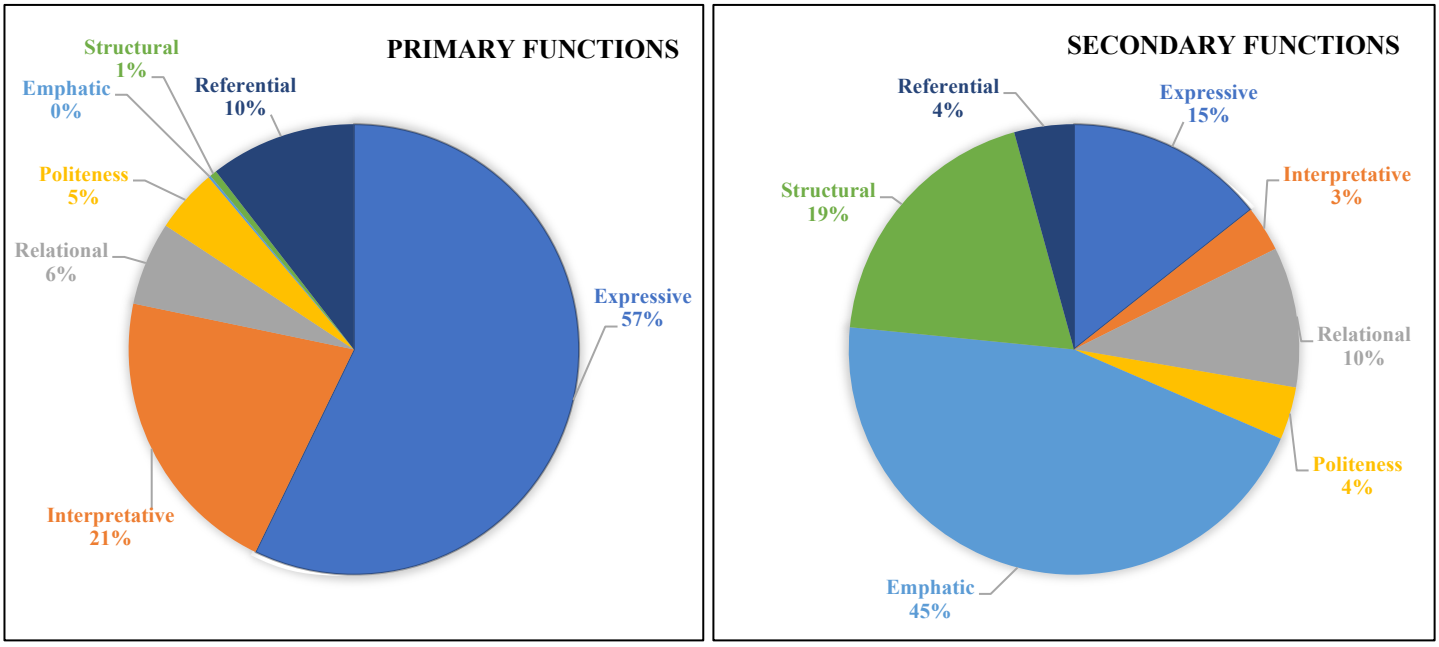
Table 9. Distribution of primary, secondary and tertiary emoji functions in the corpus.

Functions	**Functional level**					
	Primary		Secondary		Tertiary	
	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

Of the 1200 emoji, 804 were identified as having one single function, which represents 67% of the data. A secondary function was found for the remaining 33%, and a tertiary function was assigned to only 2.9% of the data (i.e. 35 emoji). The primary function of an emoji is by definition its most important function, and thus the one that should be investigated most.

Figure 1 below illustrates the distribution of the primary and secondary functions in the corpus on two pie charts. It appears clearly 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 (which represents 62.1% of all emoji). The dominance of expressive emoji is not really surprising, as conveying emotion is the most straightforward purpose of many of the most used emoji types, in particular those from the "Smileys & People" category and the "hearts" emoji.

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



The second place is occupied by the interpretative function. More than one emoji out of five in the corpus was considered to be interpretative, thus used for displaying irony or for altering the meaning of the message. Referential emoji (replacing a word or referring to a concept) represent a little more than 10% of the material. Both of these functions are almost always 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 almost never considered to be primary functions of emoji. 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.

b. Primary, secondary and tertiary functions

It is also important to look at the ways different functions combine at multiple levels for a single emoji. Table 10 shows that, in 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% of all emoji). This could be explained by suggesting that emotions are the most obvious things that can be emphasized. Therefore, the emphatic function appears frequently as the secondary function of primarily expressive emoji. Message 7 below presents an example of 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).

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

Table 10. Primary and secondary functions of the emoji in the corpus.

	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

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

It seems that the expressive function is not only the most frequent primary function (on its own or in combination with all the other functions at secondary level), but also a fruitful secondary function. Frequent combinations with the secondary expressive function include the interpretative-expressive emoji (20 cases), the relational-expressive emoji (20 cases) and the referential-expressive emoji (10 cases). A prime example of an emoji with a secondary expressive function is found in example 8. The speaker is using the "🤣" emoji to show that his sentence is ironic (as the French "Solidarity Tax on Wealth" has not actually been reestablished). The primary function of the emoji is thus the interpretative function. By repeating the emoji, the speaker is communicating that he thinks that the situation, or his own joke, is funny. Thus, emotion is conveyed: the emoji has a secondary expressive function. Additionally, the emoji is repeated three times, which strengthens the emotion and confers a tertiary emphatic dimension to the emoji.

(8) n995: "Bonne nouvelle le gouvernement rétablit l ISF pour soutenir les soignants et verser une prime aux personnes allant travailler. 🤣🤣🤣"
("Good news the government has reestablished the Solidarity Tax on Wealth in order to support the caregivers and grant a premium to the people who work. 🤣🤣🤣")

Then, the structural function appears to occur at secondary level with all types of primary functions. It is often identified in emoji that work like punctuation (e.g. between two clauses) but that have another primary function. For example, the two "🍺" emoji in message 9 are obviously 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 been needed if the emoji had not been there, in order to separate the two clauses. The emoji thus has a secondary structural function. In this case, it seems that the emoji plays the same role as a colon or a full stop.

(9) n1012: "Day out today 🍺🍺 Going hit the beer garden first,(my garden) then go have a disco,(in kitchen) then prob hit a titty bar (wife flopping them out in bedroom) night sorted buzzing for it 😍"

Tertiary functions are not presented in Table 10 above, for two reasons: firstly because they are not very frequent (only 35 emoji in the corpus have more than two functions), and secondly

because they behave almost exactly like secondary functions. Again, the expressive-emphatic combination is the most commonly found (14 cases): a secondary expressive emoji with a tertiary emphatic function. It is followed by the combination of a secondary emphatic function with a tertiary structural function (11 cases). Finally, the most frequent three-leveled combination is the following: primary interpretative, secondary expressive and tertiary emphatic functions.

c. Functions and emoji types

As seen in Table 11 below, 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 same ones (except for "🤔") also frequently have an emphatic secondary function. As outlined earlier, the expressive-emphatic combination is very common.

Table 11. The distribution of primary and secondary function for the 10 most popular emoji types (by token frequency).

	10 most frequent emoji types									
Primary function	😂	🤣	🤔	😍	😉	❤️	👍	😁	🙄	😭
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
	$X^2 = 3990.6^{***}, p < 2.2^{e-16}$									
Secondary function	😂	🤣	🤔	😍	😉	❤️	👍	😁	🙄	😭
Expressive	8	4	0	1	2	3	1	4	1	0
Interpretative	2	0	2	0	0	0	0	0	0	0
Relational	0	0	0	0	4	6	1	2	0	0
Politeness	1	0	0	0	0	0	0	1	0	0
Emphatic	26	13	0	25	0	13	9	3	2	7
Structural	5	6	4	2	3	1	3	2	3	1
Referential	0	0	0	0	0	1	0	0	0	0
	$X^2 = 2038.6^{***}, p < 2.2^{e-16}$									

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 10 below, the two emoji are primarily expressive, because they are used to represent the speaker's feeling about the movie. Then, in example 11, the emoji are used with an interpretative function. Those two most frequent emoji are very similar in appearance, but the main difference between them seems to be functional: the "🤣" emoji is more likely to be interpretative (62.5%) than the "😂" emoji (52%).

(10) n298: "Mindhorn 2017 #Netflix What a funny movie... loved it❤️🤣🤣"

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

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

Another interesting emoji is a very simple-looking one: the "winking face" emoji. Like its emoticon ancestor (";-)"), "😉" is very frequently used for displaying irony (with an interpretative function). However, it is also quite frequently used with other primary functions: the relational and the politeness functions. The emoji in example 12 and 13 illustrate this: the first "😉" is used by the speaker for maintaining a positive relationship with the addressee (relational, found 15 times in the corpus), and the second one acts as a linguistic hedge, mitigating the face-threatening value of the message (politeness, found 10 times).

(12) n948: "Bonsoir bonne lecture et bonne soirée 😉"

("Good evening, have a pleasant reading and a good night 😉"

(13) 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 😉"

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

d. Functions and categories

Not surprisingly, it appears in Table 12 that the "Smileys & People" emoji are the only ones able to be used with each of the seven functions (both at primary and secondary level). All the other emoji categories are almost exclusively referential, except for the "Symbols" emoji, which are also frequently expressive and relational (because of the high number of hearts in the corpus). As explained in section 3.1.2, some emoji in the "Animals & Nature" category can also have an emotional value.

Table 12. The distribution of primary and secondary function by emoji category.

	Emoji category							
Primary function	Smileys & People 😃	Animals & Nature 🐶	Food & Drink 🍎	Activity ⚽	Travel & Places ✈️	Objects 🛠️	Symbols ❤️	Flags 🇬🇧
Expressive	627	9	0	0	0	3	48	0
Interpretative	249	3	0	0	0	0	1	0
Relational	57	0	0	0	0	0	15	0
Politeness	54	1	0	0	0	0	0	0
Emphatic	1	0	0	0	0	0	1	0
Structural	1	0	0	0	1	0	4	0
Referential	39	15	11	4	4	10	13	30
	$X^2 = 702.3***, p < 2.2^{e-16}$							
Secondary function	Smileys & People 😃	Animals & Nature 🐶	Food & Drink 🍎	Activity ⚽	Travel & Places ✈️	Objects 🛠️	Symbols ❤️	Flags 🇬🇧
Expressive	49	1	0	0	1	0	6	0
Interpretative	12	1	0	0	0	0	0	0
Relational	31	0	0	0	0	0	8	1
Politeness	14	1	0	0	0	0	0	0
Emphatic	148	4	1	0	0	3	22	1
Structural	66	2	3	1	0	1	3	0
Referential	12	2	0	0	0	0	3	0
	$X^2 = 69.57*, p = 0.03$							

An unexpected finding concerns the "Objects" emoji, which sometimes occur with an expressive function at primary level, and with an emphatic function at secondary level (even though the relationship between category and secondary functions is not very significant, as shown by the chi-squared test in Table 12). 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 in example 14 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 the "fire" emoji, which is also frequently used to express an emotion and not to refer to proper fire.

(14) "12th feb will be a happy happy day 🎉"

e. Functions and positions

The position of an emoji appears to have a significant effect on its function. The results of chi-squared tests for independence between position and (primary and secondary) function presented below in Table 13 and Table 14 both demonstrate this high interconnection.

Table 13. 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

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

One clear finding that can be inferred from Table 12 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 word, such as the "❤️" in example 15. The emoji is used to replace the verb *love*, and fits thus right inside the clause.

(15) n615: "We ❤️ all our neighbours whatever if you're Irish, French, Dutch, Belgian, German, Spanish, Italian, Danish, Swedish, Polish, Hungarian etc..."

In Table 14, it is shown that secondarily emphatic emoji are very likely to be positioned at the end of the message in which they appear. The same observation can be made for all the

other secondary functions, except for the structural function. Structural emoji tend to appear between two clauses, which is not surprising given that the emoji placed at the intersection of two different clauses tend to play the same role as punctuation signs, and thus to play a structural role, as explained in chapter 3.1.3.

Table 14. Influence of emoji position on secondary function.

	Emoji position					
Secondary function	Final	Between	Inside	Initial	Quasi-final	Between & final
Expressive	44	5	3	2	1	1
Interpretative	12	1	0	0	0	0
Relational	38	1	0	0	1	0
Politeness	12	2	0	0	1	0
Emphatic	148	17	1	8	4	1
Structural	13	57	2	1	1	1
Referential	11	3	3	0	0	0

$X^2 = 415.12***, p < 2.2^{e-16}$

3.2.2. The influence of external factors on emoji functions

In order to answer our second research question, it is necessary to investigate whether several factors have an influence on the frequency of the different linguistic functions of emoji. This was done by computing chi-squared tests for independence between the four socio-technical variables that were selected in chapter 2 (language, platform, topic and gender) and the three levels of emoji functions. The results of these tests are presented in Table 15 below.

It appears that all four external factors have a significant influence on the primary functions of emoji. Language and topic also affect secondary functions, and platform has a small yet meaningful effect on tertiary functions. In sections *a* to *d* below, the influence of each of these factors on emoji functions are outlined and analyzed. In each section, one or two histograms are provided in order to visualize the most significant correlations of Table 15.

Table 15. Results of chi-squared tests for independence of the 4 external factors on the 3 functional levels.

	Independent factor							
Functional level	Language		Platform		Topic		Gender	
	X^2	*p*	X^2	*p*	X^2	*p*	X^2	*p*
Prim. function	**51.33*** **	<0.001	**87.27*** **	<0.001	**97.84*** **	<0.001	**79.17*** **	<0.001
Second. function	**25.69*** **	<0.001	7.87°	0.34	**52.56*** **	<0.001	17.69°	0.22
Tert. function	2.31°	0.68	**9.48***	0.05	8.18°	0.41	9.16°	0.33

a. Functions and language

A few interesting observations can be made from Figures 2 and 3, which present the distribution of primary and secondary functions of emoji across language. First, regarding primary functions, expressive emoji are more frequent in English messages, while interpretative emoji are used more by French-speaking users. What can be inferred from that is that English CMD speakers are more likely to produce emoji in order to express emotions, and that French users tend to use them to signal sarcasm more than English users. The politeness function is also more often used by French users, which could be explained by the fact that the meaning added by a politeness emoji is quite close to that of an interpretative emoji: in a way, it also alters the meaning of the message. The two functions are even gathered under one single function in a few of the previous emoji functions typologies (Cramer et al., 2106; Spina, 2018).

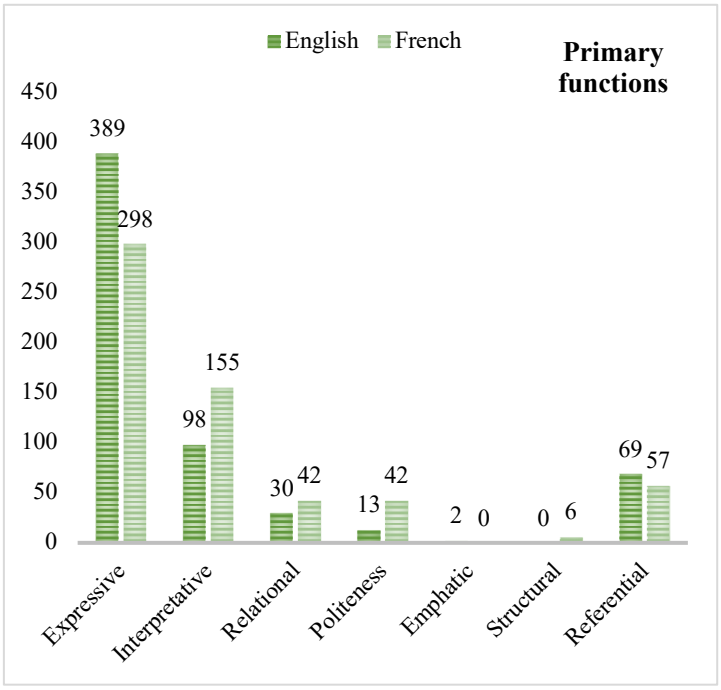


Figure 2. The influence of language on primary emoji functions.

The histogram in Figure 3, focusing on secondary functions, shows that the emphatic function is more frequent in English than in French emoji. This is not very surprising, because this function is often assigned to emoji which are primarily expressive, and it appears that primary expressive emoji are more frequently produced by English users. More surprisingly, secondary relational emoji appear three times more in English than in French, while primary relational emoji appear three times more in French.

Our results suggest that English-speaking and French-speaking CMD speakers use emoji for different reasons, and thus that language has a significant influence on the linguistic functions of emoji.

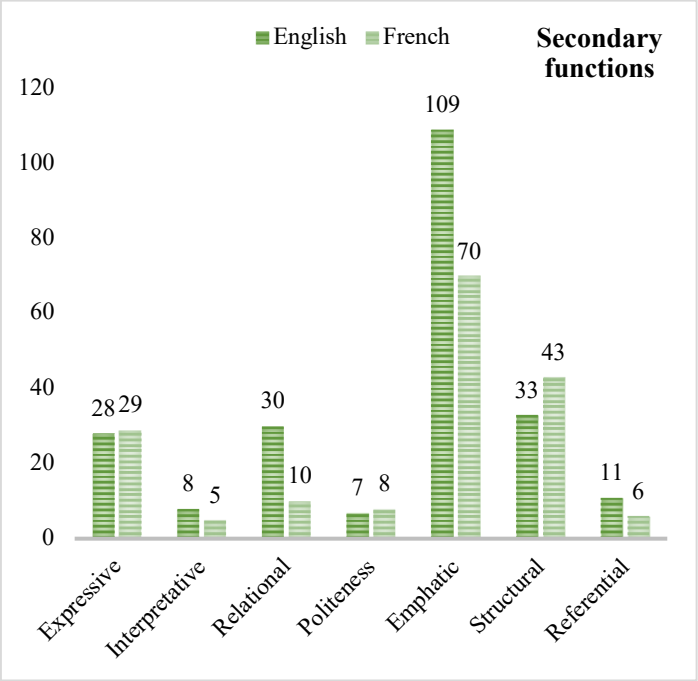


Figure 3. The influence of language on secondary emoji functions.

b. Functions and platform

There are significant differences between the primary functions of emoji produced on Facebook and those used on Twitter. First, the most striking finding from Figure 4 below is that the politeness function is almost only found in Facebook comments (94.5% of all politeness emoji were collected on Facebook). Because Facebook users are often posting under their real name (not under a pseudonym), and because they are more likely than Twitter users to know each other in real life, it is quite logical that they would be more careful with their words. Precisely, politeness emoji can be used as linguistic hedges in order to mitigate the threatening value of some utterances. Moreover, the proximity between users on Facebook can also explain the higher frequency of relational emoji in comparison with Twitter.

Then, it appears that most referential emoji were produced on Twitter (73.8% of them). Structural emoji, along with referential emoji, were also more popular on Twitter across all three functional levels (even though the effect of platform is statistically less significant for the secondary and tertiary functions, as shown earlier by Table 15). One potential reason for the

high frequency of referential emoji on Twitter is that a substantial number of Tweets in our corpus were produced by "famous" people, namely politicians and social media "influencers". In order to make their message more noticeable in the mass of Tweets posted every day, the use of emoji is a good way to attract the attention of the reader. This could be seen as a sort of "aesthetic" purpose of emoji, which will be discussed in more details in chapter 4. Referential and structural emoji can also help to make the "theme" or the "purpose" of the Tweet more rapidly understandable by the audience. For instance, the "🎬" and "🍿" emoji in example 16 both refer to the semantic field of cinema. The two referential emoji do not replace words, and the words *popcorn* and *clapperboard* are not mentioned in the Tweet: the emoji are used to announce the general theme of the message to the readers and to make this theme appear clearly with a simple glance at the Tweet.

(16) n287 & n288: "Who is ready for Netflix and chill night? 🎬🍿 Tyler Perry's A Fall From Grace streaming now on @Netflix"

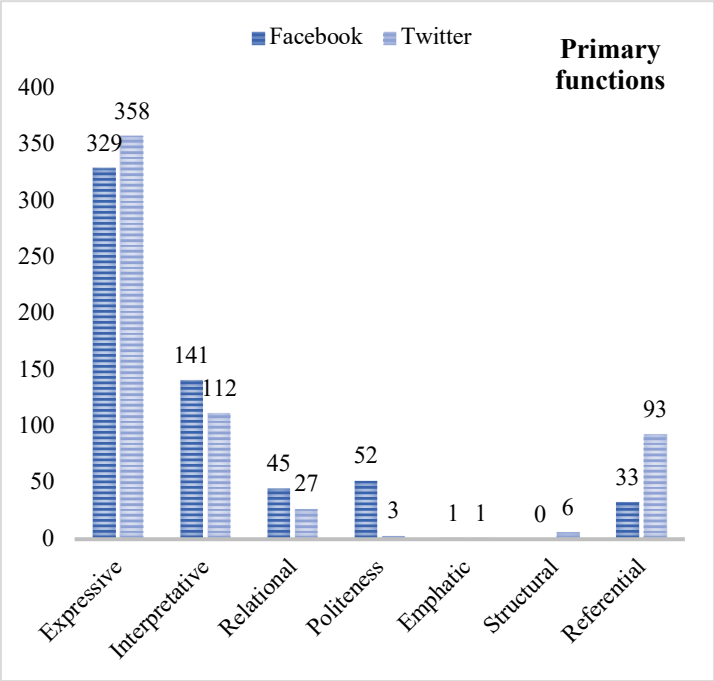


Figure 4. The influence of platform on primary emoji functions.

Lastly, expressive emoji are slightly more used on Twitter than on Facebook, while the interpretative function is more frequent on Facebook. Overall, it can be concluded that the platform on which emoji are used influences their functions, or at least that emoji are not used with the same functions on Facebook and on Twitter.

c. Functions and topic

The corpus can be divided into three subcorpora, each composed of emoji found in messages addressing three different topics: entertainment (more precisely Netflix), politics (mostly about the Brexit) and everyday life (discussions about a town and small talk on daily life during the coronavirus isolation). Figures 5 and 6 show how the frequencies of primary and secondary emoji functions differ depending on the topic of the discussion in which the emoji was used.

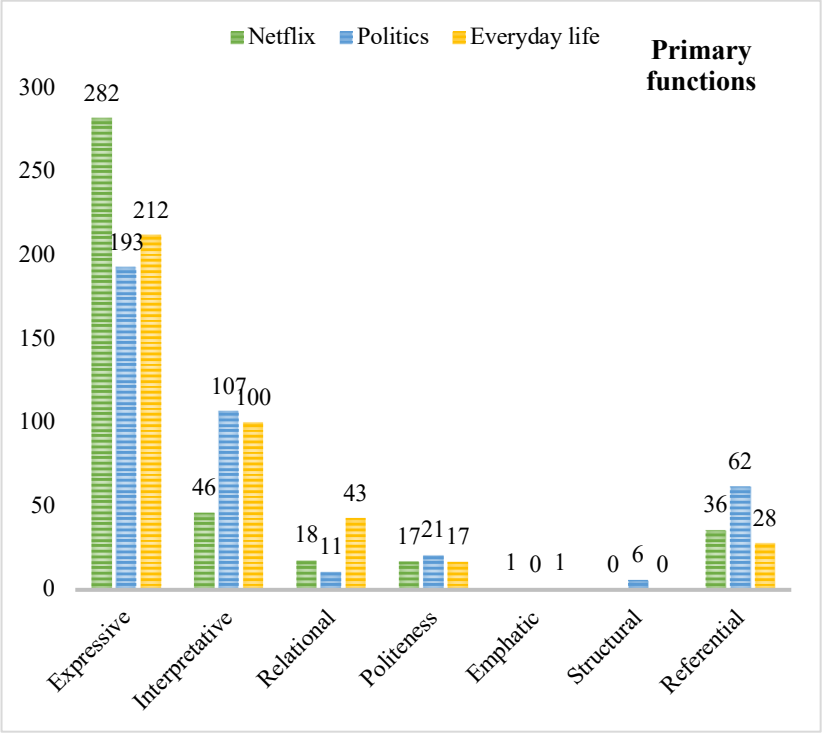


Figure 5. The influence of topic on primary emoji functions.

One striking observation that can be made from Figure 5 is that the expressive function was used much more in Tweets and Facebook comments about entertainment. This is likely due to the fact that many messages on this topic were produced in order to react to Netflix movies or shows, thus involving emotions. The three emoji in example 17 below are primarily expressive, as they are used by the speaker to show his excitement about the TV show. They are also emphatic, because their repetition reinforces the feeling of excitement (which is also conveyed by the "full caps" words and by the three exclamation marks). Therefore, the secondary emphatic function is most frequently assigned to emoji in the "entertainment" topic.

(17) n203: "SEX EDUCATION season 2 is out!!! 😍😍😍 #Netflix"

Secondly, emoji used for commenting politics are slightly less likely to be identified as primarily expressive than other emoji, as seen in Figure 5. In addition, they are very rarely used with a secondary emphatic function compared to the two other topics, as seen in Figure 6. This could be explained by the fact that people are expected to be moderate with their emotions when discussing such serious matters, which is not the case for the two other topics in the corpus (entertainment and everyday life).

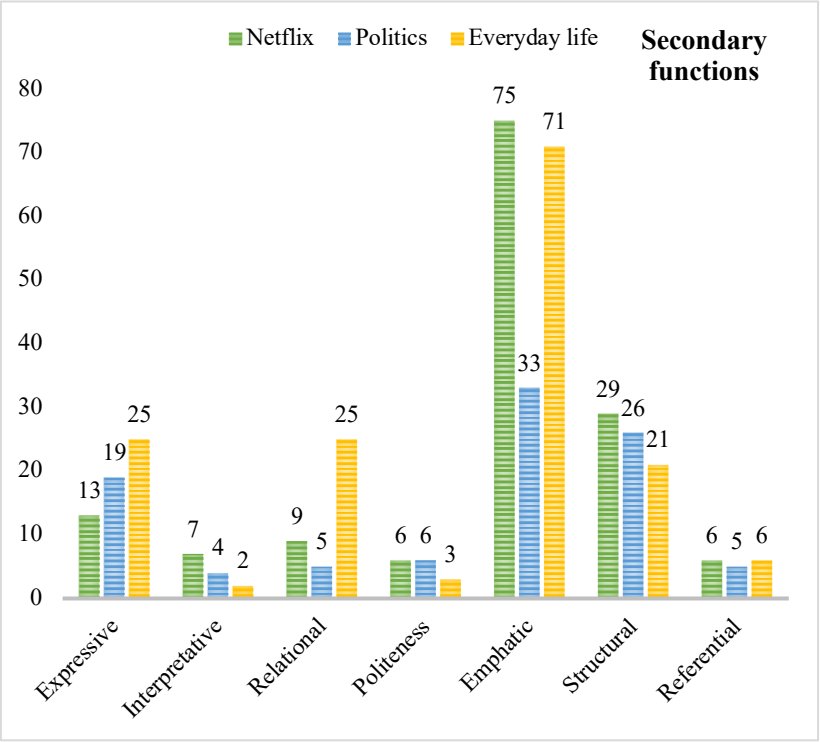


Figure 6. The influence of topic on secondary emoji functions.

Then, referential emoji are most commonly used in messages about politics, as presented in Figure 5. This relates to the fact that referential emoji are often used by politicians on Twitter in order to attract the eyes of potential readers. Another explanation for the high frequency of primarily referential emoji in politics is that most of the political discussions in the corpus were about Brexit. The flag of the United Kingdom, alongside other European flags, was therefore used very often with a referential function political messages, such as in example 18 below, in which the "🇬🇧" emoji replaces the noun *Grande-Bretagne* (*United Kingdom*).

(18) n715: “3 ans que la meute médiatique et les réseaux européistes s'obstinent à faire croire que la 🇬🇧 se mord les doigts du #Brexit.”

(“for 3 years the mediatic pack and the Europeist networks have persisted to make believe that the 🇬🇧 is kicking itself about #Brexit.”)

One last finding that can be inferred from the two histograms is that the relational function is more frequently identified with emoji about everyday life than with emoji on the two other topics, both at primary (Figure 5) and secondary level (Figure 6). A hypothetic reason for this is that many of the emoji on this topic were used as part of basic social interactions. Messages consisting of phatic formulas, such as “thank you” or “hello”, were more frequently found with this topic of discussion, and such messages were naturally often accompanied by relational emoji. The “💕” at the end of message 19 is a prime example of such relational emoji. The fact that the users interacting in these discussions know each other in real life can also be a factor influencing the higher frequency of relational emoji.

(19) n12: “So very Sad.. Rest In peace.. Condolences to his Family... 💕”

Our results suggest that the topic of discussion has a significant impact on the functions of the emoji produced by users. Topic, and more largely context of conversation, should thus be considered as a determining factor when analyzing emoji functions.

d. Functions and gender

Regarding emoji production, the main difference between men and women lies in the frequency of expressive and interpretative emoji, as can be easily inferred from Figure 7. Female users tend to produce more emoji with an expressive function, while male users appear to be more likely to add interpretative emoji in their messages. This significant difference will be addressed in more detail in chapter 4.

Even though the influence of gender on secondary and tertiary functions is not statistically significant, as shown by the chi-squared test in Table 15, it appears that emoji produced by women are more frequently identified as having a secondary emphatic function. This is probably caused by the fact that the secondary emphatic function is often associated with primary expressive emoji.

It is also surprising to observe how equally the other primary functions are distributed in terms of frequency across the two genders. The numbers of relational, politeness and referential emoji are almost exactly the same for male and female users.

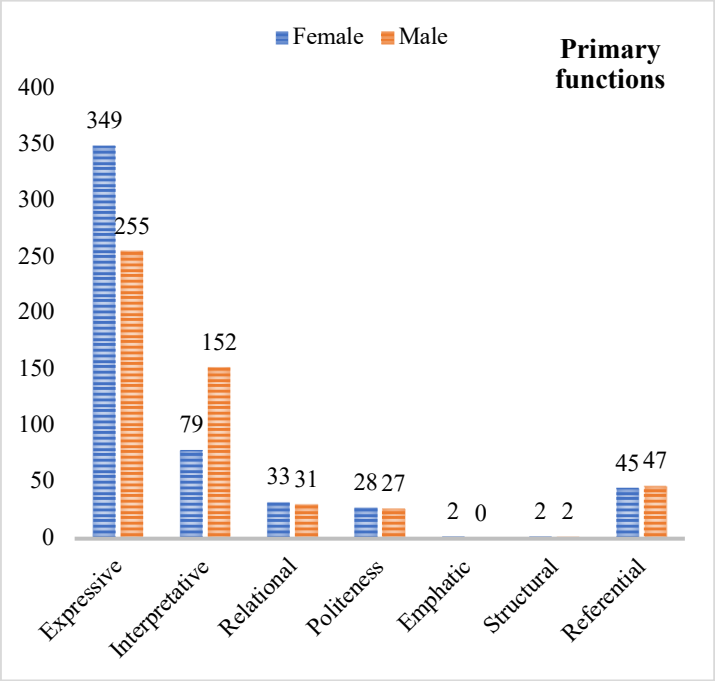


Figure 7. The influence of gender on primary emoji functions.

However, those results need to be analyzed very cautiously: as explained in section 3.1.5., gender is not normally distributed throughout the corpus. Indeed, some of the twelve subcorpora contain a majority of messages with emoji produced by female users, while others contain more emoji produced by male users. This means that the results in Figure 7 could be heavily influenced by other factors (language, platform or topic), depending on the subcorpus. Overall, there are slightly more messages produced by women than by men in the entire corpus. Some emoji were also produced by users whose gender could not be identified; these are not included in Figure 7.

In conclusion, gender has some influence on the way emoji are produced in conversations in CMD, even though this influence could be biased by other external factors. In the next section, a logistic regression analysis is computed, and its results are analyzed in order to determine whether some of the external factors are more influential than others on the emoji functions in the corpus.

3.2.3. Logistic regression analysis

In order to determine if some of the four external factors (language, platform, topic and gender) have more influence than the others on the distribution of emoji functions in the corpus, two multinomial logistic regression models were created: one for the primary emoji functions, the other one for the secondary functions. It should be noted that the emphatic and structural functions were excluded from the first model, due to the fact that they were almost never identified as primary functions.

Table 16. Model summary of a multinomial logistic regression (the four external factors applied on primary and secondary emoji functions).

	Statistical values (primary functions)			
Independent variables	*Estimate*	*Standard error*	*z*	*p*
(Intercept)	-1.250	0.15	-8.39	< 0.0001***
Gender	0.504	0.14	3.73	0.0002***
Language	0.596	0.12	4.86	< 0.0001***
Platform	-0.335	0.13	-2.57	< 0.0001***
Topic	0.817	0.16	5.25	< 0.0001***
	AIC = 1549.5			
	Statistical values (secondary functions)			
Independent variables	*Estimate*	*Standard Error*	*z*	*p*
(Intercept)	-0.329	0.14	-2.36	0.018*
Gender	-0.141	0.14	-1.02	0.309°
Language	-0.415	0.12	-3.31	0.0009***
Platform	0.001	0.13	0.01	0.99°
Topic	-0.531	0.16	-3.29	0.0009***
	AIC = 1504.5			

From the model on primary functions in Table 16, it appears that all four factors have a highly significant influence on the model. Yet, the effect of gender is slightly less significant, and platform seems to have a smaller influence on the model than the other factors. However, removing gender and language from the model does not make it significantly better, and their deletion increases the model's AIC. Therefore, it can be said that all factors have approximately the same effect, which is highly significant, on the distribution of primary emoji functions in the corpus.

The results are quite different for the logistic regression model involving secondary emoji functions. Here, the effects of both gender and platform are not significant at all inside the model. Removing the two factors from the model makes it better and causes the AIC to decrease, which means that they can be left out. However, this only applies to secondary emoji functions, which account for only a third of the data.

Overall, it appears that the external factor with the most influence on the linguistic functions of emoji is topic, followed closely by language. Gender and platform also appear to have a strong effect on the distribution of emoji functions, but only for primary functions.

4. Discussion

4.1. What are the linguistic functions of emoji?

The first section of this chapter is dedicated to our first research question, which focuses on the different linguistic functions of emoji and the typology presented in section 1.3 of this thesis. Using the statistical results described and analyzed in chapter 3 as well as some illustrative examples, we provide a quantitative and qualitative examination of each of the seven emoji functions of the typology presented in the first chapter.

After investigating every function one by one in the light of the results obtained through our corpus analysis, we will revise our initial framework in order to build, at the end of this section, an updated version of our typology of emoji functions.

4.1.1. The expressive function

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). Expressive emoji are mostly found at the end of messages, but they also occur between two clauses and in initial position. 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 "🎉").

A wide range of different emotions can be conveyed by expressive emoji. As every emoji type was designed to represent a given concept in the most immediate way, many expressive emoji types are always associated with the same specific emotion. For example, "😭" shows sadness, "😡" anger, and "👍" approval. However, some emoji appear to be more complex and are able to communicate a broader spectrum of emotions. For example, the "face with rolling eyes" emoji, "🙄", can convey not only disdain or disapproval, but also boredom.

Interestingly, it appears that emoji expressing different emotions can be combined into a single message. In example 1 below, a "❤️", showing love, is followed by two "🤣", conveying laughter. Both emoji types clearly express the speaker's emotions, as he writes that he "loved" that "funny" movie. Then, it is also possible to combine emotions which would not seem compatible at first glance, as in example 2, in which "😭" appears next to "😂". In such cases,

both expressive emoji are arranged in order to be read in a particular order, as if there was some kind of temporal dimension added to the emoji (which should be read from left to right). In example 2, the speaker shows through the emoji that the fact that he is incapable of finishing a Rubik's cube makes him sad at first, but then he laughs about it.

(1) n298: "Mindhorn 2017 #Netflix What a funny movie... loved it❤️🤣🤣"

(2) n1143: "Je crois que même avec des tutos j'en serai incapable 😭😂😂" *(talking about finishing a Rubik's cube)*
("I think that even with tutorials I would be incapable of doing it 😭😂😂")

Even though the expressive function is by far the most frequent and the easiest one to identify, some issues were raised by particular cases of expressive emoji in the corpus. First, some expressive emoji clearly appear to be used in an ironic way. For example, the "😱" in example 3 is used to express fear. The content of the message is obviously a joke in which the virus is personified as a radicalized soldier. The addition of the "face screaming in fear" emoji reinforces the absurdity of the joke. In such examples, the emoji should not be considered as interpretative, as it is not used to signal that the message is ironic. Rather, it is part of the ironical utterance. Interpretative emoji, on the contrary, are added in order to indicate that a message has to be interpreted ironically. They are not part of the ironic content of the message, as explained in section 4.1.2.

(3) n973: "Le soldat Corona 🦠 présente tous les signes apparents de radicalisation, il applique la sharria et interdit la prostitution 😱"
("Soldier Corona 🦠 shows all apparent signs of radicalization, it applies the sharia and prohibits prostitution 😱")

(4) n602: "Have you not heard about #Brexit Fabio? 🤔" *(to Fabian Zuleeg, chief executive at the European Policy Centre)*

Another prime example of what can be referred to as a "sarcastic use of expressive emoji" is found in example 4. The question asked by the speaker is clearly rhetorical and ironic. The "🤔" ("thinking face" emoji) put at the end of the message expresses that the speaker is thinking hard, and it strengthens the ironic value of that question, as the speaker is obviously not actually

trying to figure out whether European politician Fabian Zuleeg heard about Brexit or not. The emoji is expressive, but it is used sarcastically.

Then, some emoji seem to be standing on a blurred line between the expressive and the referential functions. These very particular cases occur when a speaker is describing someone's emotion with an emoji. In the Tweet presented in example 5 below, the speaker is transcribing an interview, and uses "🤷‍♀️" to represent the attitude of the interviewed person ("Woman"). Following the framework presented in section 1.3.1., this emoji could not be identified as expressive, because it does not show the emotion of the speaker of the message, but the emotion of another person. On the other hand, the four "face palm" emoji at the end of the Tweet clearly have an expressive function, as they are used by the speaker of the message to convey his own emotion towards the interview. The question is: should the "🤷‍♀️" in this example then be analyzed as a referential emoji, considering that it is used to represent a person's emotions?

(5) n700: "Woman: 'We can have a lot more rule to what happens in it (the country).' Reporter: 'Over which rules?' Woman: '🤷‍♀️, I dunno, just whatever.' 🤦‍♀️🤦‍♂️🤦‍♀️🤦‍♂️"

(6) n670: "Taxi driver this morning:" 'Did you have a #Brexit party last night? That EU, they tried to put all their laws on us but they don't do it themselves. Like, on holiday, you can smoke in all the bars, but they don't let us here.'
Me: Looks out of window 😶"

The same problem is illustrated in example 6, in which the speaker is reporting a conversation between a taxi driver and herself. An emoji, "😶", is used once again to represent the emotion (here, embarrassment) of one of the participants of the conversation. However, this time, this participant is the speaker herself. Therefore, it could be argued that this emoji is expressive, as it is used to convey the speaker's own emotion.

In order to solve this issue, it might be necessary to broaden the definition of expressive emoji when reviewing our functional typology later in this chapter. Expressive emoji should be considered as able to convey not only the speaker's emotions, but also the emotions of other individuals mentioned by the speaker.

Overall, the expressive function is unequivocally the most important linguistic function of emoji in context. It can be used in various situations and can be assigned to many of the most popular emoji types. Therefore, it should keep its predominant place inside our revised typology of emoji functions.

4.1.2. The interpretative function

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 in 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. Yet, they are also often used with the expressive function.

Some messages may contain two or more emoji which all appear to have an interpretative function. In example 7, the content of the message is ironic, and the first "🤣" emoji signals the irony to the reader. However, it is unclear how the two other emoji should be interpreted. First, they could be analyzed as also interpretative, as well as emphatic, because they are used to insist on the fact that the message is ironic. But they can also be seen as expressive, showing that the speaker considers his own ironic message to be funny. Example 8 could help solving this problem. The first of the three "😠" radically modifies the tone of the message, and signals that it has to be understood ironically. The next two emoji should not be considered as two more indicators of irony, as they would be redundant. Actually, they show that the speaker is angry about the situation, and are thus expressive emoji. Therefore, the new framework should specify that, when several emoji occur in an ironic message, only one emoji (usually the first one) carries the interpretative function.

(7) n635: "#Brexit Britain sounds like an absolute HOOT! 🤣😂🤣"

(8) n636: ""Ne pas se précipiter" pour protéger les français ! Ils ont raison on a tout notre temps l'épidémie ne progresse pas 😠😠😠"

("Not rushing" in order to protect the French! They're right, we have all the time, the epidemic is not progressing 😠😠😠")

By definition, interpretative emoji guide the interpretation of text. Yet, it seems that they are also able to alter the meaning of other emoji. For instance, the interpretative emoji "😛" in example 9 is used to show that the content of the message is ironic, including the expressive

emoji "☺️". This expressive emoji is used sarcastically (the speaker does not actually like the fact that people insult him because he's gay), and the interpretative emoji signals that.

(9) n397: "Dans ces moments là, j'aime quand vous me balancez que je finirais en enfer parce que je suis homo ☺️😛"

("It these moments, I like it when you tell me that I'll end up in hell because I'm gay ☺️😛")

Then, some emoji in the corpus seem 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 section 1.3.2. These particular cases suggest that the definition of the interpretative function may need to be extended. First, in example 10, the "smirking face" emoji, "😏", which is almost always used with an interpretative function in the corpus, is placed at the end of the message. However, the content of the message is not ironic; 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 example 10, the speaker's implicature is: "what Miller says does not make sense". The role of "😏" is to make the reader aware of that implicature, and not to signal irony or to alter the tone of the message. In a similar way, the "🤑" emoji in example 11 influences the reader's interpretation of the message, showing that the question is rhetorical and that the speaker implies that "Netflix just wants to make money, so they definitely won't choose France for the taxes". In other words, in this case, the emoji signals that the message contains an implicature, and also guides the reader's interpretation towards the implicature.

(10) n738: "Donc pour #Miller les Anglais ont quitté l'Europe de #Macron ...en 2016 vote du #Brexit et #Macron a été élu en 2017 ..#NonRien 😎😏" (*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.. #NoNothing 😎😏")

(11) n308: "Et pour les impôts ? Est-ce que #Netflix va également #ChooseFrance? 🤔🤑"

("And for the taxes? Will #Netflix also #ChooseFrance? 🤔🤑")

In Grice's account, ironical utterances can also be analyzed as implicatures, where "what is implicated" is the opposite of "what is said" by the speaker (Clark, 2013). For this reason, it could 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.

4.1.3. The relational function

Relational emoji were not found very frequently 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. As for the expressive function, the relational function is often produced by facial emoji, gesture emoji, and hearts. 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. In example 12, the "💕" clearly has a relational function, but it also conveys a positive emotion expressed by the speaker towards the situation or the addressee. Similarly, the three "🥰" in example 13 can be analyzed as relational, because they are used by the speaker to build a positive relationship with her readers. However, it is obvious that these emoji are also communicating a positive emotion towards the readers, which is even emphasized by their repetition. Therefore, they have a tertiary emphatic function as well.

(12) "I went looking for them but they'd all gone xx is this a regular thing to look out for? 💕" *(talking about chocolate brownies in a bakery)*

(13) "Boooonnjooour à vous tous les gens et les ami.e.s twittos 🥰🥰🥰 #Confinementjour12 et demi...."
("Goooood afternoon to all the people and the Twitter friends 🥰🥰🥰 #Isolationday12 and a half....")

Because the relational function appears to be intrinsically linked to the expressive function, it seems important to review the position occupied by this function inside the typology. The relational function could be considered as a subordinate function, only occurring with emoji already analyzed as expressive.

4.1.4. The politeness function

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, as explained in section 1.3.4. They constitute small "positive" softeners in generally "negative" messages.

Following the framework introduced by Brown & Levinson (1987), politeness emoji can be used to soften utterances in which the face of the speaker is potentially menaced, such as apologies (as in example 14). Yet, 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 15.

(14) n167: "Sorry je n'ai pas lu, je voulais juste dire que c'est en cours 🙂"
("Sorry I haven't read, just wanted to say that it's ongoing 🙂")

(15) 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. 😉")

Sometimes, two or more emoji are used with the politeness function in the same message. For instance, in example 16, the "🤓" used by the speaker to soften the criticism is produced three times. It could be said that this repetition adds an emphatic value to the politeness emoji, and thereby an even stronger softening effect on the criticism. Two different emoji types with the politeness function can also be combined, as in example 17. Here, the first politeness emoji

is "😇". The second politeness emoji, "😘", can be considered as having a secondary emphatic function, additionally to its primary politeness function, as it is used to weaken even more the potential threat in the message (a criticism).

(16) n942: "Pour le moment la seule victime de votre dernier post c'est l'orthographe 🤓🤓🤓"

("For now the only victim of your last post is the spelling 🤓🤓🤓")

(17) n523: "même si en France, on aime pas trop le changement, c'était toujours mieux avant, bah là c'était vraiment mieux avant, ça fait moins pro, moins Netflix 😇😘" *(commenting on Netflix's new interface)*

("even though in France, we don't really like change, it was better before, well here it was actually better before, it looks less professional, less Netflix 😇😘"

Some of the messages in the corpus were part of heated conversations that could sometimes result in insults, mostly when discussing sensitive matters (e.g. politics). Emoji were also used with a politeness function in order to diminish the highly threatening value of these insults. The "face blowing a kiss" emoji following the insult in example 18 below illustrates this. In some sense, it can be said that politeness emoji such as this one are used to avoid – or simply to weaken – impoliteness.

(18) n695: "Happy #Brexit Mothercluckers 😘"

(19) n675: "@JohnHannah is certainly very angry and clearly not looking to bring the country 🇬🇧 together

Also how can he work again as he will constantly be reminded of these comments Silly, silly, silly. 🤡🤡🤡" *(John Hannah is a British actor who stands against Brexit)*

(20) n784: "Encore un démocrate qui refuse les choix du peuple ! 🤡"

(21) n1115: "Et cette merde voulait être maire de #Bordeaux 💩💩💩"

("and this piece of shit wanted to be mayor of #Bordeaux 💩💩💩")

Then, another remarkable finding is that emoji can not only be used to soften insults; they can also function independently as "weak" insults. This is shown in examples 19 and 20: in both messages, the "clown face" emoji are used to call someone a clown, without actually saying the word "clown". The insult is not mentioned with words, but with emoji, which

function as a weaker form of the same insult. The colorful and humorous value of emoji are used to mitigate the threat in the original insult. Therefore, these "emoji insults" can be considered as having a politeness function. However, they are also in some sense referential, because they are used as referents of a particular concept. The referential and politeness functions are not incompatible, as attested by the three "💩" in example 21. These emoji refer to the very offensive word "merde" in the message and are used to soften it. They can thus be considered as having both a referential and a politeness function.

Even though it does not appear very frequently in the corpus, the politeness function is still used by many CMD speakers and has to find its right place into our typology. Yet, it is unclear whether this function should be considered as one of the main emoji functions or as a subordinate function. One could argue that all politeness emoji are essentially interpretative emoji, because they are used to alter the meaning of the message, but considering that they can work as insults, their role appears to be more complex.

4.1.5. The emphatic function

The first important observation to make is that emoji were only assigned the emphatic function as a secondary function, and it was almost only assigned to primarily expressive emoji. As explained in section 3.2.1., 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, as in example 22. The first "angry face" emoji is repeated seven times, which strongly emphasizes its emotional value.

(22) n982: "Ahhhhh mais attends....ils prennent le temps ! 😠😠😠 ça se voit qu'ils sont bien à l'abri dans leurs bureaux ces planqués ! 😤😤😤😠😠😠😠😠😠"
("Ahhhhh but wait... they're taking their time! 😠😠😠 it's clear that they're well sheltered in their offices, these privileged people! 😤😤😤😠😠😠😠😠😠"

Then, emoji combination can be another way to add an emphatic function to expressive emoji. In the same example above, three "😤" are combined with the eight "😠", insisting even more on the anger conveyed by the speaker. Similarly, the "thumbs up" emoji in example 23 is used to emphasize the positive emotion already expressed by the "😊". Still, although it has a secondary emphatic function, the primary role of the "👍" is to convey a positive emotion, and

it is thus mainly expressive. A third possibility is that emoji can be used to emphasize an emotion that is already expressed in the textual content of the message (and not by another emoji). This is illustrated by example 24, in which the "😞" is used to reinforce the feeling of sadness already communicated by the adverb "sadly".

(23) n219: "Some good times had in there! 😊👍"

(24) n20: "my aunty and uncle sadly passed 😞"

As discussed in section 4.1.2. above, interpretative and relational emoji seem unable to carry a secondary emphatic function. On the other hand, the politeness function can be emphasized by other emoji or through emoji repetition, as seen in section 4.1.4. But there is another important combination that has not been discussed yet: the referential-emphatic combination. Indeed, in some messages in the corpus, emoji with a clear referential function are repeated two or more times. For instance, the "🇬🇧" in example 25 is used to refer to the fact that the Netflix show "The Bodyguard" is British. However, it is unclear why this emoji appears three times in a row: how could reference possibly be emphasized? Therefore, it could be suggested that some referential emoji can also carry an expressive function, and that their repetition can subsequently cause the emergence of an emphatic function. This will be further investigated in section 4.1.7., in which the referential function is discussed in more detail.

(25) n242: "If you like British crime then The Bodyguard on #Netflix is absolutely bloody fantastic!!! 🇬🇧🇬🇧🇬🇧"

These different observations lead to the conclusion that the emphatic function should definitely be reconsidered as a subordinate function in the revised version of our typology, as it is always attached to emoji with another main function: expressive, politeness or (maybe) referential emoji.

4.1.6. The structural function

From our results, 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), as explained in section 3.2.1. The structural function is almost always secondary, as the emoji separating clauses often have another primary function. A good illustration of this is the "red heart" emoji in example 26 below. The emoji clearly has an expressive function, but it is also used to separate the speaker's main clause from her final comment ("missing our Scrabble Thursday's").

(26) n1011: "Today my 82 year old grandma used Facebook video chat for the first time and it made my day❤️ missing our Scrabble Thursday's."

(27) n980: "Analyser pourquoi nous n avons pas de matériel aussi ❓ ❓ ❓ 😕 😕 😕" ("Analyzing why we do not have material as well ❓ ❓ ❓ 😕 😕 😕")

(28) n709: "🔴 Je suis aujourd'hui au Parlement européen de Bruxelles, avec notre groupe « Identité & Démocratie », qui rassemble désormais 10 nationalités et devient le 4ème force parlementaire suite au #Brexit !"
("🔴 Today I am at the European Parliament of Brussels, with our group « Identity & Democracy », which now gathers 10 nationalities and becomes the fourth parliamentary power due to the #Brexit!")

However, for some emoji of the corpus, it could be argued that the structural function is the primary function, or even the only function. In example 27 above, three " ❓ " emoji are used to replace regular interrogation marks. Yet, the fact that they are emoji, along with their repetition, adds an emphatic aspect to the expression of confusion, which is further strengthened by the combination with the three "😕" emoji. Thus, the " ❓ " are essentially expressive and emphatic, and the structural function could therefore be considered as tertiary in this message. Example 28 is more complicated. The "large red circle" emoji in initial position seems to be simply used to introduce the Tweet. The use of "Symbols" emoji for this purpose, and particularly of the "🔴" emoji, appears to have become a norm for news accounts and political profiles on Twitter, in order to signal that something new or important is being reported. It is thus used, in some sense, to communicate the "theme" of the message to the reader, which could be considered as a referential use. For this reason, in this example, the structural function is also combined with another primary function, the referential function. Finally, the "🔴" is also used to attract the attention of the reader, which could be related to a sort of "aesthetic function". Again, this will be examined in more detail in the next section of this chapter.

Another interesting structural emoji is found in example 29. In this message, the speaker is first quoting Emmanuel Macron, and then using the expressive emoji "🙄" to show his disapproval. Then, he uses two "musical notes" emoji in order to introduce and frame his comment (i.e. "Le problème, c'est larem"). These two "🎶", one before and one after the comment, function like parentheses, they "frame" the clause, which gives them a structural function. But they are also used to show that this final comment has to be understood as a slogan or a catchphrase (as *problème* rhymes with *larem*). The "🎶" are used to evoke the musicality of this motto and can therefore be considered as both referential and structural.

(29) n707: "Selon Emmanuel Macron, l'Europe, "c'est le choix de la Nation française, plusieurs fois réaffirmé dans des référendums" 🙄 🎶 Le problème, c'est larem 🎶" *("LaREM", or "La République En Marche", is a French political party led by Macron)*
("According to Emmanuel Macron, Europe is "the choice of the French Nation, multiple times reasserted in referenda" 🙄 🎶 The problem is larem 🎶")

Like the emphatic function, it is clear that the structural function should not be analyzed as a primary emoji function in our revised typology. It works as a supplementary or subordinate function that all emoji types can carry and that can be added to all the other functions.

4.1.7. The referential function

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 not able to convey something else than the iconic concept they refer to. For example, the "🍎" emoji is typically only used to refer to the concept "apple". Still, as was already explained in chapter 3, there are some exceptions (e.g "🎉", "🔥" or "🙈").

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, like the two "Flags" emoji in example 30, which are produced instead of the words *the United Kingdom* and *the European Union*. Adjectives can also be replaced by emoji, as with the "European flag" emoji standing for the word *Européenne* in example 31. Word replacement by

referential emoji also appears to be possible for adverbs or adverbial phrases, like the "💯" in example 32, and for verbs, like the "🗳️" in example 33.

(30) n619: "The rule for #Brexit talks from now on: it's *always* easier for 🇬🇧 to just blame 🇪🇺 for anything and everything, whether justified or not, than it is to face up to the costs and trade offs inherent in Brexit."

(31) n755: "Ce jour marquera le début de la fin de l'utopie 🇪🇺, celle qui a cru possible la négation de 1500 ans d'Histoire des nations."
("Today will mark the beginning of the end of the 🇪🇺 utopia, which though that the negation of 1500 of History of nations would be possible.")

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

(33) n766: "Le 21 Mars 2021 ils 🗳️" *(ils votent)*
("On 21 March 2021 they 🗳️") *(they vote)*

Some referential emoji replace words, but others can also refer to words or concepts that are already present in the verbal content of the message, such as the "🍷" in example 34 below. Two or more emoji can also be combined in order to refer to a precise concept, as is illustrated by example 35, in which the "pizza" emoji are combined with the "seedling" emoji to refer to the "vegan pizza" concept. Then, referential emoji can also evoke a concept that is not directly mentioned in the message but that adds some contextual information to it. For instance, the "🐓" in example 36 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. Another illustration of this is found in example 37, in which the three "🐑" help the reader deriving the implicature that "Western stock market investors" are implicitly compared to sheep. The "🦠" emoji in example 3 is also used by the speaker to explain his own joke by disambiguating the fact that he is referring to the coronavirus. Based on these observations, it appears that the referential function is not incompatible with the interpretative function.

(34) n1056: "Blossom from last nights walk. Having a mid afternoon glass of wine 🍷...not like I'll be driving anywhere ☹️"

(35) n1037: "Vegan pizza! 🍕🌱🍕🌱🍕🌱"

(36) n71: “Wednesday nights 👌🐓”

(37) n901: “Predictably, the covid19 pandemic has led to panic-selling of shares in first class key Western companies, as western stock market investors dive for cover before you can say ‘Baaa... 🐑🐑🐑’.”

(3) n973: “Le soldat Corona 🦠 présente tous les signes apparents de radicalisation, il applique la sharria et interdit la prostitution 😱”

Emoji can refer not only to words or concepts inside the message, but also to elements outside it. For example, the “⬇️” in example 38 refers directly to a video located below the message. Likewise, the “👇” in example 39 draws the attention of the reader to the article attached to the Tweet. Therefore, these two emoji can also be said to play a deictic role, which is part of their referential function. The “⬇️” is even accompanied by the deictic pronoun “this”, referring to the video below the Tweet. Then, the two “😂” in example 40 are used to refer to the content of a picture on which the speaker is commenting. This is a good example of how facial emoji can also have a referential function. Another particular facial emoji is the “🤓” in example 41. In this message, the emoji is used by the speaker, a famous French TV star, to refer to himself iconically, as he is known for always wearing the same characteristic pair of glasses. Aside from its referential function, this emoji can also be analyzed as both expressive and relational, in this context. It should also be noted that this emoji can only be understood as referential by someone who knows this TV presenter’s face, and thus requires some cultural knowledge. This will be further discussed in section 4.3.

(38) n694: “Whenever you feel sad, thinking you had a bad day .... watch this ⬇️” *(with a video making fun of anti-Trump politicians)*

(39) n612: “Great. Let's ✂️ red tape to help businesses expand and our economy grow. 👇” *(with a link to a news article on the economic benefits of Brexit)*

(40) n947: “Un bon dessin vaut mieux qu'un long discours 😉😂😂” *(with a comic explaining that people should stay home during the coronavirus crisis)*
(“A good picture speaks a thousand words 😉😂😂”)

(41) n1131: “Reposez-vous bien… et à demain 🤓” *(by a French TV presenter who is easily recognizable by his pair of round glasses)*
(“Rest well... and see you tomorrow 🤓”)

As suggested in the two previous sections, it appears that emoji which are used to refer to a concept are able to have more than just a referential function. Firstly, as seen in example 25 above (section 4.1.5.), referential emoji are sometimes repeated, even though one emoji would technically be sufficient to evoke the concept referred to. Example 42 below illustrates this as well: the first "🏎" emoji is used by the speaker to refer to the TV series "Drive to Survive", yet it is unclear why this emoji is repeated. A possible explanation for this is that referential emoji can also carry an expressive function, and that their repetition adds emphasis to the emotion expressed. The emotions that can be conveyed by referential emoji appear to be generally positive, as in example 25 and 42, or even in example 43. The three "🎥" seem to reinforce the excitement already expressed by the interjection ("Oh") and the verbal content of the message ("très bonne nouvelle").

(42) n441: "F1 Drive to Survive S2 out on my birthday. Get In 🏎🏎"

(43) "Oh la très bonne nouvelle 🎥🎥🎥 A voir en ce moment #TheIrishman, superbe ambiance et jeux d'acteurs incroyables"

("Oh the great news 🎥🎥🎥 Watch #TheIrishman now, great atmosphere and incredible acting")

Secondly, it can be argued that emoji, in particular referential emoji, add some sort of aesthetic value to the message in which they are produced. As explained in section 1.3.8., emoji can be used with a decorative purpose, making the message more colorful and more attractive to the eyes of 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.

It would not be reasonable to say that all emoji are produced with an aesthetic purpose. However, in some messages, it is unclear what the actual purpose of the emoji is, other than the addition of "decoration" to the message. This is often the case with referential emoji, such as those in examples 30 to 35 discussed earlier, or the "🎨" in example 44 below, which does not seem to add anything more than a piece of ornament added to the message.

(44) n1044: "These are all sold out now but not before I prepared this (very pleased with my artistic skills here 🎨)."

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. One of these functions has been identified as being the aesthetic function, which should be added into our revised typology as a subordinate emoji function.

4.1.8. A new typology of emoji functions

After a thorough analysis of our quantitative results and a qualitative examination of some interesting examples found in the corpus, it appears that our first hypothesis (H1), which is repeated below, is partially correct. Indeed, the initial typology of emoji function presented in the section 1.3. is operational, but it requires improvement.

- H1: The "7-functions typology" needs improvement: some functions will be proven unnecessary after the corpus analysis, and new important functions will emerge.

The first and most important adjustment that has to be made to the typology is to distinguish between 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 appeared 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 the referential functions), while others were always seen as secondary (i.e. the emphatic and the structural functions). Therefore, in this new typology, it was decided to identify, from the start, essential primary and essential secondary functions. The five secondary functions can only be assigned to emoji which already have one of the three primary functions (expressive, interpretative of referential). For example, an emoji can only be classified as being relational if it has an expressive function. However, this does not exclude that an emoji can have two primary functions, or two secondary functions, simultaneously: for example, an emoji can be both referential and expressive, as in example 41 in section 4.1.7.

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 1 below.

Our revised typology of emoji functions arose from the analysis of a corpus composed of 1200 emoji in context. These emoji were found in French and English Tweets and Facebook comments on a variety of topics. The typology has been proved to be applicable in 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 is also operational in other similar contexts of conversation and in other languages, though this hypothesis would have to be verified through further research.

Table 17. Revised typology of the linguistic functions of emoji in conversation.

Primary functions	Expressive	Conveying emotion, whether the speaker's own emotional state or someone else's.
	Interpretative	Signaling the presence of an implicature in the message and guiding the interpretation of the implicature (including ironical and sarcastic utterances).
	Referential	Referring to a concept inside or outside the message, by replacement, evocation or repetition.
Secondary functions	Relational	Maintaining a positive relationship with the reader(s) of the message. Combined with expressive and referential emoji.
	Politeness	Softening the potential face-threatening value of a message. Combined with expressive and interpretative emoji (or referential emoji for weakened insults).
	Emphatic	Strengthening the emotional value of a message or an emoji. Combined with expressive emoji (and politeness emoji).
	Structural	Separating, framing and punctuating clauses. Combined with expressive, interpretative and referential emoji.
	Aesthetic	Adding a decorative value to the emoji, making the message more attractive, more colorful or playful. Mainly combined with referential emoji.

4.2. What influences the functions of emoji?

In order to answer our second research question, we collected data divided into twelve different subcorpora, allowing us to compare the frequencies of emoji functions efficiently depending on four external factors: platform, language, topic and gender. The analysis of age as a factor of influence had to be abandoned due to technical limitations, as explained in section 2.2.1.

At the end of our examination, it appears that all four factors have a significant influence on the linguistic functions of emoji in context, which confirms our hypothesis. In the four following sections, the influence of each external factor is presented and discussed.

4.2.1. Platform

Our investigation focused on two different social networking sites, Facebook and Twitter. Facebook is the most popular SNS in the world, and Twitter has already been investigated by many researchers. The main similarities and differences between the two platforms were described in section 2.2.2. Still, it should be pointed out that our research only focused on emoji used in messages that were shared publicly or posted on public and semi-public pages and groups. When producing the messages and the emoji, the users were aware that they could potentially be read by thousands of people or more. The results of our study would probably have been much different if the corpus included emoji used in private contexts, like the instant messaging apps of Facebook and Twitter.

When looking at the differences between discourse on Facebook and on Twitter, Lin & Qiu (2013) found that Twitter users tend to show less verbal intimacy in their conversations than Facebook users. These observations reflect the fact that users tend to communicate with people from their real-life close relationships (i.e. family, friends, colleagues) on Facebook, and with people outside these circles on Twitter. Our findings appear to confirm this trend. As outlined by the results detailed in section 3.2.2., Facebook users are more likely to use emoji with a relational function than Twitter users, and are much more likely to use politeness emoji. These two functions are assigned to emoji which are used to have a positive impact on the relationship between users, which is likely linked to the higher intimacy between Facebook users.

Then, Lin & Qiu (2013) also suggested that Facebook users tend to show their emotions through non-verbal cues (i.e. emoticons and emoji) more often than Twitter users. However, our results show that Facebook users are actually less likely to use emoji with an expressive

function than Twitter users. The referential function is also a lot more frequent on Twitter, as detailed in section 3.2.2.

Generally, our results have made clear that emoji functions are not produced with the same frequencies on Facebook and on Twitter. Therefore, it can be argued that the platform on which emoji are produced has an influence on the linguistic functions of these emoji. It can be expected that this difference would be even greater between emoji used on social networking sites and emoji used in instant messaging. This would have to be examined in another study.

4.2.2. Language

The two languages explored in this study are French and English. Most French emoji on Facebook were collected on groups containing mainly Belgian users, thus speaking Belgian French. Most English emoji on Facebook were gathered from groups containing mainly British users, thus speaking British English. However, on Twitter, it would have been too complicated to pick only messages written by users speaking specifically these two regional varieties. Therefore, our final results cannot be considered as accounting for Belgian French and British English only, but for French and English in general. It would be interesting to pursue future research by analyzing whether emoji functions vary from a specific variety to another of the same language (e.g. British English and American English).

A previous study on the differences of emoji production across languages and countries showed that French and English were not significantly different regarding the types of emoji that they used most (Lu et al., 2016). Yet, they found that French users tend to produce more emoji than other users, and that they use much more "heart" emoji than users of other parts of the world. Although it is not possible to verify the first observation with our results, as our corpus counts the same amount of French and English emoji, the second observation is surprisingly not confirmed by our data: English-speaking users are more likely to use hearts than French-speaking users. The contradiction between our results and the previous ones (Lu et al., 2016) could be explained by the fact that our study compares languages, while the other study focused on countries. This difference suggests even more that cultural and regional factors, along with language, can have a significant influence on the linguistic functions of emoji.

As shown by our results (see section 3.2.2.), language has a significant impact on the frequencies of emoji functions. Messages written in English are more likely to include emoji used for conveying emotions (expressive and emphatic emoji), while emoji produced by

French-speaking users are more often interpretative. The reason for this contrast is unknown, but it could be suggested that French-speaking users are more likely to communicate their emotions through verbal content than English-speaking users, who prefer to express them with emoji. This would require a deeper investigation. To conclude, it can be said that our results show that language has an influence on the reasons why people use emoji. Although we only studied French and English emoji, two languages that are quite close to each other, it is likely that the functions of emoji would vary across other languages, even more when comparing languages from different linguistic families or culturally different regions of the world, as suggested by previous studies (Barbieri et al., 2016a; Lu et al., 2016).

4.2.3. Topic

The topic of discussion is one of the first social factors highlighted by Susan Herring in her classification scheme for CMD genres (Herring, 2007). Several studies showed the importance of the context and topic of conversation on emoji production (Cramer et al., 2016; Beißwenger & Pappert, 2019).

As shown by the linear regression models computed and described in section 3.2.3., the most influential factor (when all four external factors are considered simultaneously) is the topic of conversation. The three different topics chosen for our study, namely entertainment, politics and everyday life, appear to cause different frequencies of emoji functions. These results are not surprising, as detailed in the third chapter, because they reflect how people would be expected to behave in face-to-face conversations about these same subjects. Some interesting observations include, first, that expressive and emphatic emoji are very likely to be found in messages about entertainment, whereas they are much less frequent in political discussions. Then, when discussing more intimate topics (everyday life), users tend to produce more relational emoji.

It has to be noted that the three topics chosen for this study do not obviously represent all the different contexts of conversation in which emoji are likely to appear. It can be expected that emoji used in messages commenting on other topics that have not been investigated here would display different frequencies of linguistic functions. Yet, the three topics studied in this thesis were selected because they represent very different conversational contexts, which could explain the strong differences found through our results.

4.2.4. Gender

Of all four external factors analyzed in this study, gender is the one that has been investigated the most by previous research on emoji variation. A recent study conducted in China concluded that gender had no significant influence on emoji production (An et al., 2018). Our results, on the other hand, seem to show that a user's gender has some effect on the linguistic functions of the emoji produced by that user.

When investigating emoticons, Wolf (2000) discovered that women produce them more frequently than men. This finding was confirmed by later studies on emoji, which found that they were produced more by female than by male users (Chen et al., 2017; Jones et al., 2020). Even though our research did not focus on the total frequency of emoji production depending on gender, slightly more emoji in our corpus were produced by women (538 emoji, against 514 by men). This difference is not significant, but it seems to support previous observations. Then, in our corpus, it appears that "hearts" emoji are more frequently used by women than by men. This totally contradicts previous findings (Chen et al., 2017), but this difference could be explained by the fact that our study is based on public messages, whereas the 2017 results included emoji used in private conversations. Indeed, it can be suggested that users behave differently in private and in public discussions, as social norms and pressure are in force on social networking sites but not in intimate conversations.

Regarding emoji functions, Wolf (2000) found that female users are more likely to use emoji in order to express their emotions than male users, while men tend to produce more interpretative emoji than women. These findings are confirmed by our results, as seen in section 3.2.2. Both Wolf (2000) and Alshenqeeti (2016) admit that this finding appears to strengthen the stereotype of the emotional women and the humorous men. Because our data was collected from public conversations, it can be suggested that the male and female users were acting accordingly to the social roles and stereotypes that have been allocated to them, and that these results could be significantly different with private conversations. In the end, even though the results only differ between men and women regarding the frequencies of expressive and interpretative emoji, it is clear that gender has a significant influence on the linguistic functions of emoji.

4.3. Limitations

Even though our analysis led to many very interesting findings and discussions, there are some inherent limitations to our study which have to be taken into consideration.

First, the most important of these limitations concerns the fact 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 the functions identified by the annotator. Analyzing every emoji one by one required a lot of time, because it had to be done very carefully in order to avoid misinterpretations. Therefore, when it appeared to be too difficult to identify an emoji's linguistic function (or functions), that particular emoji was excluded from the corpus. 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. In the end, although the risks of misinterpretation were minimized, it should be noted that it is never possible to know exactly what function was intended by a user when producing an emoji (except by asking them directly), and that some emoji in the corpus may thus have been misinterpreted.

It is also possible that some emoji were misused by the speakers. In example 45 below, for instance, it is unclear why "🙄" is added to the message, other than for an expressive purpose. The "face with rolling eyes" emoji is usually used to convey disdain, frustration or sarcasm, according to Emojipedia.[14] Yet the message is part of a friendly conversation in which the speaker is not showing any sign of hostility towards his addressee. The only plausible explanation is that the emoji is the consequence of a typo, or of a confusion due to a lack of experience with emoji. An expressive emoji showing thought such as "🤔", or a relational emoji such as "😊", would have probably been more appropriate in the context of this message. This suggests that "emoji misuse" may have occurred in other places in the corpus, and that some emoji may have been wrongly interpreted because of that. On the other hand, as in verbal language, misuse can be seen as a part of authentic language use.

(45) n95: "Great old picture. I remember the cannons, I think? Were they near the old boys grammar school? 🙄"

[14] https://emojipedia.org/face-with-rolling-eyes/, last accessed on May 30, 2020.

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. As explained in section 1.2.1., the appearances of emoji can be different 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 have the ability to predict and suggest emoji depending 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.[15] 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 thesis 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, in order to verify how much the linguistic functions of emoji are influenced by the time period during which they are produced.

[15] https://theconversation.com/coronavirus-emojis-conveying-compassion-and-humour-with-a-facemask-137322, last accessed on May 30, 2020.

Conclusion

Throughout the different chapters of this thesis, the main objective was to understand how and why people use emoji in conversation. After a summary of previous studies on the subject, a multimodal corpus was built and analyzed in detail, quantitatively and qualitatively, in order to answer two research questions.

The first one aimed to verify whether the initial typology of emoji functions, which was presented in section 1.3. and composed on the basis of earlier frameworks and personal observations, was operational. In the end, it appeared that this "7-functions" framework was partly successful in classifying the 1200 emoji of the corpus. However, some interesting examples of emoji in context showed that the typology had to be reworked: an eighth function, the aesthetic function, was added, and the framework distinguished primary and secondary functions. Even though this new typology was based on observations made from our corpus, it would be relevant to test its efficiency on another corpus of emoji in context. This could be achieved by a future research project.

The goal of the second research question was to determine if the linguistic functions of emoji in context are influenced by four external factors: the CMD platform, the language, the topic of conversation and the gender of the speaker. The speaker's age was also initially considered to be included in the analysis but had to be excluded due to technical limitations.

In sum, it appears that all four factors have a significant influence on emoji functions. First, two social networking sites were compared. It was found that the politeness and relational functions are more frequently used on Facebook, and that referential emoji are more common on Twitter. Regarding language, English users are more likely to produce emoji to convey emotions, whereas French speakers use them more frequently to disambiguate ironical utterances or to soften potentially face-threatening utterances. However, these results could be related to cultural rather than linguistic differences, but this would have to be investigated in further research. Topic is also very influential, as emoji used in conversations about entertainment are often used to show strong emotions, while in political discussions they are more commonly produced with the referential function. Relational emoji are much more frequent in small talk and conversations about everyday life. Then, it appears that female users are more likely than male users to use emoji to show their emotions. Male users, on the other hand, use interpretative emoji more often than their female counterparts.

These results were obtained using the initial "7-functions" typology, it would thus be interesting to examine the results of a similar analysis made on the basis of our new typology

of emoji functions. It can be suggested that the results would be quite similar to those found in this study, yet they would be more accurate, as the revised typology is certainly more operational than the initial one.

In further research, it would also be interesting to compare the differences in emoji functions on other CMD platforms, in other languages and on other topics than those analyzed in our study. For example, comparing emoji functions in private messaging apps and on social media, or between linguistically and culturally more different languages (e.g. English, Arabic and Japanese), would surely yield valuable results. Then, focusing on the influence of other factors on the variation of emoji functions would also be a path worth investigating. Time period, age and culture were already mentioned as factors of interest for future research. Another interesting variable that could influence the linguistic functions of emoji is the proficiency of users not only with CMD, but also with the language itself. Does the linguistic level of speakers influence how they use emoji? Do native speakers and learners use the same emoji functions? And more, does the L1 of a speaker influence its use of emoji when communicating in his or her L2? These very interesting questions could be answered in a future research project.

Finally, it should be noted that the results of our study may have implications for other purposes than research. First, the typology could be used by "smart keyboards" software companies, such as Swiftkey, in order to improve the quality of emoji prediction. For example, typical politeness emoji could be automatically suggested to users when their message appears to be potentially face-threatening. Similarly, our findings on the influence of platform, language, topic and gender could be taken into account in order to propose the right emoji to the right users and in appropriate contexts. Then, in the field of natural language processing (NLP), "chatbots" could also be enhanced by knowing when to add specific emoji with the adequate functions in particular contexts. This way, our results could help making conversations with some of these virtual assistants appear a lot more natural.

Bibliography

Adamic, L., Develin, M. & Weinsberg, U. (2015). The Not-So-Universal Language of Laughter. Facebook research. Retrieved October 5, 2019, from https://research.fb.com/the-not-souniversal-language-of-laughter/.

Alden, J. (2019). Hedging rule discussions. A study on hedging and emoticons in an online board game discussion forum. *Bachelor degree's thesis*. Falun: Högskolan Dalarna.

Alhabash, S., & Ma, M. (2017). A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Social Media + Society, 3.1*, 1–13.

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.

An, J., Li, T., Teng, Y., & Zhang, P. (2018). Factors Influencing Emoji Usage in Smartphone Mediated Communications. *International Conference on Information*, 423-428.

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics 10(4)*, 419-438.

Aull, B. (2019). A study of phatic emoji use in WhatsApp communication. *Internet Pragmatics*. John Benjamins Publishing Company, p. 1-34.

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

Barbieri, F., Kruszewski, G., Ronzano, F. (2016a). How Cosmopolitan are emojis? Exploring emojis over meaning over different languages with distributional semantics. *Proceedings of the 2016 ACM on Multimedia Conference*, 531-535.

Barbieri, F., Espinosa-Anke, L., & Saggion, H. (2016b). Revealing patterns of Twitter emoji usage in Barcelona and Madrid. *Frontiers in Artificial Intelligence and Applications 288*, 239-244.

Barbieri, F. & Camacho-Collados, J. (2018). How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. *Proceedings of the 7th Joint Conference on Lexical and Computational Semantics (*SEM 2018)*, 101-106.

Barbieri, F., Camacho-Collados, J., Ronzano, F., Anke, L. E., Ballesteros, M., Basile, V., Patti, V. & Saggion, H. (2018). Semeval 2018 task 2: Multilingual emoji prediction. *Proceedings of The 12th International Workshop on Semantic Evaluation*, 24-33.

Baron, N. (1984). Computer mediated communication as a force in language change. *Visible Language*, *18.2*, 118-41.

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

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

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

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.

Blagdon, J. (2013, March 4). How emoji conquered the world. *The Verge*. Retrieved 19 October 2019 from https://www.theverge.com/2013/3/4/3966140/how-emoji-conquered-the-world.
Brisson, C.M. (2015). *Hieroglyphs at Our Fingertips*. [Online]. Retrieved 20 October 2019 from https://www.academia.edu/12568625/Hieroglyphs_at_Our_Fingertips_Language_Semiotics_and_Communication_through_Emoji.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness. Some universals in language usage*. Cambridge: University Press.

Buccafurri, F., Lax, G., Nicolazzo, S., & Nocera, A. (2015). Comparing Twitter and Facebook user behavior: Privacy and other aspects. *Computers in Human Behavior, 52*, 87-95.

Carr, C. T., & Hayes, R. A. (2015). Social media: Defining, developing, and divining. *Atlantic Journal of Communication, 23*, 46–65.

Cavalcanti, L. H., Pinto, A., Brubaker, J. R. & Dombrowski, L. S., (2017). Media, Meaning, and Context Loss in Ephemeral Communication Platforms: A Qualitative Investigation on Snapchat. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*, 1934–1945.

Chan, M. (2011). Shyness, sociability, and the role of media synchronicity in the use of computer-mediated communication for interpersonal communication. *Asian Journal of Social Psychology, 14*, 84–90.

Chanier, T., Poudat, C., Sagot, B., Antoniadis, G., Wigham, C. R., Hriba, L., Longhi, J., Seddah, D., (2014). The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres. *Journal of Language Technology and Computational Linguistics* 29.2, pp. 1-30.

Chen, Z., Lu, X., Shen, S., Ai, W., Liu, X., & Mei, Q. (2017). *Through a gender lens: An empirical study of emoji usage over large-scale android users*, Peking University, 1-10.

Clark, B. (2013). *Relevance Theory*. Cambridge: Cambridge University Press.

Cougnon, L.-A. (2015). The Not-So-Universal Language of Laughter: French results. Retrieved April 20, 2019 from http://cental.fltr.ucl.ac.be/team/acougnon/lol.

Cramer, H., de Juan, P. & Tetreault, J. (2016). Sender-Intended Functions of Emojis in US Messaging. P*roceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 504–509.

Crible, L. (2017). Towards an operational category of discourse markers: A definition and its model. In Fedriani, C. & Sansó, A. (Eds.), *Discourse markers, pragmatics markers and modal particles: New perspectives*. Amsterdam: John Benjamins.

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

Dimson, Thomas. 2015. *Emojineering Part 1: Machine Learning for Emoji Trends*. Instagram Engineering Blog. Retrieved 19 October from https://instagram-engineering.com/emojineering-part-1-machine-learning-for-emoji-trendsmachine-learning-for-emoji-trends-7f5f9cb979ad.

Dresner, E. & Herring, S. (2010), Functions of the Non-Verbal in CMC: Emoticons and Illocutionnary Force, in *Communication Theory*, 249-268.

Ellison, N.B. & Boyd, D.M. (2013). Sociality through social network sites. In Dutton, W.H. (Ed.). *The Oxford Handbook of Internet Studies.* Oxford: Oxford University Press, 151–172.

Escouflaire L. (2018a), *Quelles représentations linguistiques sur l'âge et le genre d'une personne son utilisation de smileys en CMO véhicule-t-elle ?* Bachelor thesis. Under the supervision of Hambye, P., UCLouvain.

Escouflaire, L. (2018b). *"Haha" as a discourse marker in French and English CMD*. Under the supervision of Crible, L. & Degand, L. CLIG2230: Discourse Analysis. UCLouvain.

Escouflaire, L. (2018c). *L'influence de l'âge et du genre d'un locuteur sur son utilisation de smileys en CMO francophone.* Under the supervision of Dister, A. CLIG2250: Méthodologie de l'analyse de corpus en linguistique, UCLouvain.

Fleiss, J. L. and Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, Vol. 33*, 613–619.

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.

Gallud, J. A., Fardoun, H. M., Andres, F., & Safa, N. (2018). A Study on How Older People Use Emojis. *Proceedings of the XIX International Conference on Human Computer Interaction*, 1-4.

Gawne, L. & McCulloch, G. (2019). Emoji as digital gestures. *Language@ Internet 17*.

Goffman, E. (1967). *Interaction ritual. Essays on face-to-face behavior*. Garden City/New York: Doubleday.

Goldmark, A. (2014, March 19). Inside the Mind That Translated Moby Dick into Emoji. *New Tech City*. Retrieved 27 November 2019 from https://www.wnyc.org/story/emoji-moby-dick-translation-fred-benenson/.

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

Gumperz, J. (1982). *Discourse strategies*. Cambridge, England: CUP.

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

Hall, J. A., Horgan, T. G., & Murphy, N. A. (2019). Nonverbal communication. *Annual Review of Psychology, 70*, 271–294.

Herring, S. C. (Ed.). (1996). *Computer-mediated communication: Linguistic, social and cross-cultural perspectives.* Amsterdam: John Benjamins.

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet, 4.1.*

Herring, S. C. (2013). Discourse in Web 2.0: Familiar, reconfigured, and emergent. In Tannen, D. & Tester, A. M. (Eds.), *Georgetown University Round Table on languages and linguistics 2011: Discourse 2.0: Language and new media*. Washington, DC: Georgetown University Press, 1-25.

Herring, S. C. (2019). The co-evolution of computer-mediated communication and computer-mediated discourse analysis. In Bou-Franch, P. & Garcés-Conejos Blitvich, P. (Ed.). *Analysing digital discourse: New insights and future directions*. London, Palgrave Macmillan, 25-67.

Herring, S. C. & Androutsopoulos, J. (2015). Computer-mediated discourse 2.0. In Tannen, D., Hamilton, H. E. & Schiffrin, D (Ed.). *The handbook of discourse analysis (2nd Ed.)*. West Sussex: Wiley Blackwell, 127-151.

Herring, S. C., Dainas, A. (2017). "Nice picture comment!" Graphicons in Facebook comment threads, *Proceedings of the 50th Hawaii international conference on system sciences.* Waikoloa, Hawaii, 2185–2194.

Herring, S. C., & Dainas, A. R. (2018). Receiver interpretations of emoji functions: A gender perspective. In *Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji2018). Stanford, CA*.

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 Humanitie*, 30-33.

Hoyle, R., Das, S., Kapadia, A., Lee, A. J. & Vaniea, K. (2017). Was my message read? Privacy and signaling on Facebook messenger. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '17),* 3838–3842.

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109.19, 7241-7244.

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.

Jones, L. L., Wurm, L. H., Norville, G. A. & Mullins, K. L. (2020). Sex differences in emoji use, familiarity, and valence. *Computers in Human Behavior, 108*.

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.*

Larsson, A. O. (2017). Diversifying Likes. *Journalism Practice*, 12.3, 326-343.

Lebduska, L. (2014). Emoji, emoji, what for art thou? *Harlot: A Revealing Look Arts Persuasion*, *1.12*.

Lin, H. & Qiu, L. (2013). Two sites, two voices: Linguistic differences between Facebook status updates and tweets. In Rau, P. L. P. (Ed.). *Lecture notes in computer science, 8024 LNCS.* Berlin: Springer, 432–440.

Ljubešić, N., & Fišer, D. (2016). A global analysis of emoji usage. *Proceedings of the 10th Web as Corpus Workshop*, 82-89.

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.

Maíz-Arévalo, C. (2015). Typographic alteration in formal computer-mediated communication. *Procedia, 212*, 140–145.

Malik, M. M., Lamba, H., Nakos, C. (2015). Population bias in geotagged tweets. *Ninth International AAAI Conference on Weblogs and Social Media*, 18–27.

Mäntymäki, M., & Islam, A. K. M. N. (2016). The Janus face of Facebook: Positive and negative sides of social networking site use. *Computers in Human Behavior, 61*, 14–26.

Marcoccia, M. (2000). La représentation du non verbal dans la communication écrite médiatisée par ordinateur. *Communication et organisation, 18*, 265-274.

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.

Mellon, J. & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics, 4.3*.

Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L. & Hecht, B. (2016). "Blissfully happy" or "ready to fight": Varying interpretations of emoji. *Proceedings of ICWSM 2016*, 259-268.

Miller, H., Levonian, Z., Kluver, D., Terveen, L., & Hecht, B. (2018). What I see is what you don't get: The effects of (not) seeing emoji rendering differences across platforms. *Proceedings of the ACM on Human-Computer Interaction 2*/*124*, 1-21.

Mislove, A., Lehmann, S., Ahn, Y. Y. (2011). Understanding the demographics of Twitter users. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 554–57.

Moschini, I. (2016). The" Face with Tears of Joy" Emoji. A Socio-Semiotic and Multimodal Insight into a Japan-America Mash-Up. *HERMES-Journal of Language and Communication in Business, 55*, p.11-25.

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*, p. 136–141.

Nishimura, Y. (2015). A sociolinguistic analysis of emoticon usage in Japanese blogs: Variation by age, gender, and topic. *AoIR Selected Papers of Internet Research*, *5*.

Novak, P. K., Smailović, J., Sluban, B. & Mozetič, I. (2015). *Sentiment of Emojis*. PloS One 10 (12): e0144296.

Ortiz-Ospina, E. (2019, September 18). *The rise of social media*. Our World in Data. University of Oxford. Retrieved October 14 from https://ourworldindata.org/rise-of-social-media.

Panckhurst, R. (2006). Le discours électronique médié : bilan et perspectives. dans Piolat, A. (dir.). *Lire, Écrire, communiquer et apprendre avec Internet*, Marseille, Editions Solal, 345-365.

Page, R., Barton, D., Unger, J. W. & Zappavigna, M. (2014). *Researching Language in Social Media*. London: Routledge.

Park, J., Barash, V., Fink, C., & Cha, M. (2013). Emoticon style: Interpreting differences in emoticons across cultures. *Seventh International AAAI Conference on Weblogs and Social Media,* 466-475.

Park, J., Baek, Y. M., & Cha, M. (2014). Cross-cultural comparison of nonverbal cues in emoticons on twitter: Evidence from big data analysis. *Journal of communication*, 64.2, 333-354.

Pavalanathan, U., & Eisenstein, J. (2015). Emoticons vs. emojis on Twitter: A causal inference approach. arXiv preprint arXiv:1510.08480.

Petitjean, C. & Morel, E. (2017). ”Hahaha”: Laughter as a resource to manage WhatsApp conversations. *Journal of Pragmatics, 110*, 1–19.

Pohl, H., Stanke, D., & Rohs, M. (2016). EmojiZoom: emoji entry via large overview maps. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 510-517.

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.

Quito, A. (2019, October 18). *Why we can't stop using the "face with tears of joy" emoji*. Quartz. Retrieved 24 October 2019 from https://qz.com/1726756/the-psychology-behind-the-most-popular-emoji/.

R Core Team (2019). *R: A Language and envionment for statistical computing*. (Version 3.6) [Computer software]. Retrieved from https://cran.r-project.org/.

Raclaw, J. (2006). Punctuation as Social Action: The Ellipsis as a Discourse Marker in Computer-Mediated Communication. *Proceedings of the 32nd annual Berkeley Linguistics Society, 32.1,* 299-306.

Radford, W., Chisholm, A., Hachey, B., & Han, B. (2016). Telephone. Person. Sailboat. Whale. Okhand or 'Call me Ishmael' — How do you translate emoji? *Proceedings of Australasian Language Technology Association Workshop*, 150–154.

Rafi, M. S. (2008). SMS text analysis: Language, gender and current practices. *Online Journal of TESOL France*, 1-13.

Rodrigues, D., Lopes, D., Prada, M., Thompson, D. & Garrido, M. V. (2017). A frown emoji can be worth a thousand words: perceptions of emoji use in text messages exchanged between romantic partners. *Telematics and Informatics, 34.8*, 1532-1543.

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.

Shoeb, A. A. M., Raji, S. & De Melo, G. (2019). EmoTag–Towards an Emotion-Based Analysis of Emojis. *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 1094-1103.

Simons, G. & Fennig, C. (2019). *Ethnologue*, *22nd edition*. Dallas, TX: SIL International.

Skovholt, K., Grønning, A. & Kankaanranta, A. (2014). The Communicative Functions of Emoticons in Workplace E-Mails: :-). *Journal of Computer-Mediated Communication 19*, 780–797.

Smith, A. (2011). *22% of online Americans used social networking or Twitter for politics in 2010 campaign.* Washington, DC: Pew Research Center, 1-92.

Spina, S. (2018). Role of Emoticons as Structural Markers in Twitter Interactions. *Discourse Processes, 56.4*, p. 345-362.

Stark, L. & Crawford, K. (2015). The conservatism of emoji: Work, affect, and communication. *Social Media + Society, 1.2, 1-11.*

Steinmetz, K. (2015, November 16). Oxford's 2015 Word of the Year Is This Emoji. *Time*. Retrieved 19 October 2019 from https://time.com/4114886/oxford-word-of-the-year-2015-emoji/.

Tagliamonte, S. A. & Denis, D. (2008). Linguistic ruin? Lol! Instant messaging and teen language. *American Speech 83.1*, 3-34.

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.

Tauch, C., & Kanjo, E. (2016). The roles of emojis in mobile phone notifications. *Proceedings of the 2016 acm international joint conference on pervasive and ubiquitous computing: Adjunct*, 1560-1565.

The jamovi project (2020). *jamovi*. (Version 1.2) [Computer Software]. Retrieved from https://www.jamovi.org.

Tossell, C. C., Kortum, P., Shepard, C., Barg-Walkow, L. H., Rahmati, A., & Zhong, L. (2012). A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior*, *28.2,* 659-663.

Troiano, G. & Nante, N. (2018). Emoji: What does the scientific literature say about them? A new way to communicate in the 21th century. *Journal of Human Behavior in the Social Environment*, 28(4), 528-533.

Tuttle, M. (2016). *Emoji essay. Emoji: Imperfectly Filling a Gap in Text-based Communication*. Retrieved 19 October 2019 from https://blogs.ubc.ca/mollytuttle/essay/.

Unicode (2019, October 18). *Emoji frequency. The most frequent emoji*. Retrieved 24 October 2019 from https://home.unicode.org/emoji/emoji-frequency/.

Uygur-Distexhe, D. (2012). Lol, mdr and ptdr. An inclusive and gradual approach to discourse markers. *Linguisticae Investigationes 35.2*, 389-413.

Weissman, B. (2019). Peaches and eggplants or… something else? The role of context in emoji interpretations. *Proceedings of the Linguistic Society of America, 4.29*, 1-6.

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).*

Wilberdin, K. & Wells G. (2019, February 4). Facebook's Timeline: 15 Years In. *The Wall Street Journal*. Retrieved October 11 from https://www.wsj.com/articles/facebooks-timeline-15-years-in-11549276201#comments_sector.

Wiseman, S. & Gould, S. S. J. (2018). Repurposing Emoji for Personalised Communication: Why the pizza emoji means "I love you". *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. *Article 152,* 1-10.

Yus, Francisco (2014): Not all emoticons are created equal. *Linguagem em (Dis)curso 14.3*, 511–529.

Zanotti, E. (2019, October 2). The 'Sexy Peach' Emoji Is The Newest Symbol Of The Resistance. *The Daily Wire.* Retrieved 8 November 2019 from https://www.dailywire.com/news/the-sexy-peach-emoji-is-the-newest-symbol-of-the-resistance.

Zhao, P., Jia, J., An, Y., Liang, J., Xie, L., & Luo, J. (2018). Analyzing and predicting emoji usages in social media. *Companion Proceedings of the The Web Conference 2018*, 327-334.

Zipf, G. (1935). *The psychobiology of language*. Houghton Miffin.