



## From Emojis to Sentiment Analysis

Gaël Guibon, Magalie Ochs, Patrice Bellot

### ► To cite this version:

Gaël Guibon, Magalie Ochs, Patrice Bellot. From Emojis to Sentiment Analysis. WACAI 2016, Lab-STICC; ENIB; LITIS, Jun 2016, Brest, France. hal-01529708

**HAL Id: hal-01529708**

**<https://hal-amu.archives-ouvertes.fr/hal-01529708>**

Submitted on 31 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# From Emojis to Sentiment Analysis

Gaël Guibon  
Aix Marseille Université,  
CNRS, ENSAM, Université de  
Toulon, LSIS UMR7296,13397  
Caléa Solutions  
Marseille  
France  
gael.guibon@lsis.org

Magalie Ochs  
Aix Marseille Université,  
CNRS, ENSAM, Université de  
Toulon, LSIS UMR7296,13397  
Marseille  
France  
magalie.ochs@lsis.org

Patrice Bellot  
Aix Marseille Université,  
CNRS, ENSAM, Université de  
Toulon, LSIS UMR7296,13397  
Marseille  
France  
patrice.bellot@lsis.org

## ABSTRACT

Studies on Twitter are becoming quite common these years. Even so, the majority of them did not focused on emoticons, even less on emojis. An overview of emoticons related work has been made recently [11]. However there is still too little research work related to emojis. In this paper we draw up the work and future approaches worth considering for emoji usage in Sentiment Analysis. We aim to put necessary theoretical background before using emojis for sentiment analysis. Thus, we present an emoji usage typology along with linguistic and socio-linguistic studies on the interpretation of emojis. We also introduce approaches exploiting emojis in Sentiment Analysis. We conclude by presenting our perspectives in this domain considering the evolution of emoji usages.

## Keywords

emoji; emoticon; sentiment analysis; natural language processing

## 1. INTRODUCTION

Emojis are an unavoidable emerging data of this last years, especially for marketing, digital communication but also for information retrieval dedicated to sentiment analysis and opinion mining. In a recent report the Emoji Research Team of Emogi<sup>1</sup> has shown that 92% of the online population are using emojis [25]. However, while the majority of sentiment analysis works in Natural Language Processing (NLP) uses Twitter, which contains emojis and emoticons, only a few focuses on the role of emoticons for sentiment analysis, even less about emojis. In this paper we make an overview of several works done in the field of sentiment analysis exploiting emojis. Indeed, sentiment analysis studies specialized on emojis are scattered. Thus, our main objective is to draw up the different approaches which could be relevant to the usage of emojis for sentiment analysis purposes. Most of the works on sentiment analysis were focused on limited numbers of emojis and did not consider the extent of the sentiment representation they convey. In this paper our objectives are to try to represent the extent of the representation of the sentiment of emojis, and to make an overview of the works exploiting emojis for sentiment analysis, or the ones which could be relevant for it. By doing so, we intend to put the theoretical background related to the exploitation of emojis

for sentiment analysis, to highlight the different possibilities they offer. [12]

The paper is organised as follow: At first, after defining emoticons and emojis (Section 2), we present a categorization of the usages of emojis before exploring the impact it have on the emoji's interpretation (Section 3). Then, we try to find which computer science approaches could be relevant to emojis' usage for sentiment analysis (Section 4). Finally, we propose the perspectives of an approach which could be useful considering all the previous aspects (Section 5).

## 2. DEFINITIONS

In this section, we first define emoticons and emojis separately, as they are often amalgamated. Emojis being sometimes considered as simple traduction of emoticons [1].

### 2.1 Emoticons

To define emoticons we make the distinction between two types of emoticons. This distinction is rather unofficial (meaning categorized by users<sup>2</sup>) but old enough<sup>3</sup> and separates emoticons into two types: *western emoticons*, and *eastern emoticons*. This distinction is important as most of sentiment analysis works do not consider *eastern emoticons* but only *western emoticons*.

The term emoticon is a shortcut for "emotion icon". Emoticons were considered as ASCII characters sequences [21], a succession of characters, and not an image. They usually are meant to represent only faces and facial expressions and were first used by professor Scott Fahlman in 1982 [15]. However, sometimes emoticons are used to represent objects (Figure 1) or body parts (i.e o/ -Head with an arm up-, oTZ -a man bowing- ).

**Western emoticons.** They are the most commonly used and known emoticons. They usually are horizontal and have a limited representativeness ( = ), xD , etc ).

**Eastern emoticons.** Also known as "Japanese emoticons" or *kaemoji*, in this paper we will only refer to them as eastern emoticons. Eastern emoticons are vertical and can represent more complex faces and body positions than western emoticons. Such as o(0\_0)/ (surprised with an arm up), /(Q.Q)\(crying, sad with arms downside), or Figure 1 showing that emoticons can also represent objects, actions, and emotions.

### 2.2 Emojis

<sup>2</sup>[https://en.wikipedia.org/wiki/List\\_of\\_emoticons](https://en.wikipedia.org/wiki/List_of_emoticons)

<sup>3</sup><http://marshall.freeshell.org/smileys.html>

<sup>1</sup><http://emogi.com/>





**Figure 1: Objects in emoticons: Angry man throwing a table**

*Emoji* is a Japanese word meaning “picture characters”. By opposition to emoticons, emojis are not characters but images. They were first created by Shigetaka Kurita in the late 90s using Unicode characters. They were meant to be an evolution of eastern emoticons and to represent not only faces, but also concepts, objects, etc. They first appeared on the DoCoMo mobile operator brand<sup>4</sup>, but were in fact popularized worldwide by Apple with the iPhone which included them. As shown in Table 3, emojis can represent different kind of things, and as they are pictures, they can also be combined into one single emoji.

	Simple	Combined
<i>bust in silhouette emoji</i>		
<i>woman with bunny ears emoji</i>		

**Table 1: Example of a combined emoji**

Emojis were dependent of one software architecture at first. Meaning that an emoji was represented with a unique code from the software company, and that the picture associated to the code could also be intern to the company. Since their usage rises up, the most common emojis have been standardised by Unicode<sup>5</sup>. The standardisation concerns only the code, thus different graphical representations of emojis emerged. For instance, Twitter has focused on flat design compared to other companies (flat design chestnut , non flat design chestnut ). These different representations may lead to different interpretations. Miller [16] made a survey to quantify those differences. For instance, in Table 2 same emojis, depending on the platform where it is displayed, may convey different emotional or cognitive states. The unicodes associated to the emojis are then not sufficient to express emotional or cognitive states. The image itself, and more particularly the emotional facial cues determined both by the Unicode and the platform, should be considered.

## 2.3 Emojis and emoticons

Emojis enable one to represent by pictures the equivalent of several emoticons. However, some emoticons (as illustrated in Table 4) has no equivalent in terms of emojis, and vice versa. Moreover, the most used emojis are not only made of “faces emojis”<sup>6</sup>, at least on Twitter. Emoticons tend to slightly disappear, and especially western emoticons. This

<sup>4</sup><https://www.nttdocomo.co.jp/english/index.html?param>

<sup>5</sup>The standardized unicode emojis can be found here:<http://unicode.org/emoji/charts/full-emoji-list.html>


<sup>6</sup><http://emojitracker.com/>

	Apple	Samsung
Face with rolling eyes emoji		
Polarity Emotions	Slightly Negative Unhappy / Sad / Doubtful	Positive Happy / Surprised

**Table 2: face with rolling eyes emoji: Opposite emotions and polarity**

Faces	Body actions	Objects	Ideas	Places

**Table 3: Emojis can represent many things...**

can be explained by the fact that standard emoticons (e.g. :-)) can be easily replaced by emojis (). However, eastern emoticons offer more possibility in terms of customization, and as such they keep being used. In fact, where emoticons are customizable, emojis are already done. This feature reinforced the use of eastern emoticons as a customizable alternative to emojis, with a quite intensive usage in forums or chat rooms. Even more for really famous forums such as 4chan<sup>7</sup> or Reddit<sup>8</sup>, where specific emoticons can act as a community identifier. The latter forum even supported the creation of a website dedicated to eastern emoticons<sup>9</sup>. The imagination and satisfaction for a person or a group to use and possess their own type of emoticons cannot be done with out-of-the-box emojis. Moreover, now emoticons are not only made with ASCII characters anymore, as Unicode characters<sup>10</sup> are used in order to represent something extending again the customization possibilities.

Consequently, we must focus on emojis and we consider the emoticons that do not have emoji representations. Moreover, in this paper we distinguish emoticons from emojis. But they can be close to each other under several circumstances, that is why even if we focus on emojis, we also consider emoticons.

In the next section we present the different types of usage for emojis, and the impact they could have on their interpretation by one or multiple addressees.

## 3. USAGE AND INTERPRETATION OF EMOJIS

From a semiotic point of view, emojis and emoticons are mere signs, and as such, they possess a meaning<sup>11</sup> [22]. To understand a symbol, a sign, it is important to take into account the context of its appearance.

### 3.1 Usages







<sup>7</sup><http://www.4chan.org/>

<sup>8</sup><https://www.reddit.com/>

<sup>9</sup><http://japaneseemoticons.me/>

<sup>10</sup>Someunicodeemoticonscanbefoundhere:<http://unicodeemoticons.com/>

<sup>11</sup>Peirce theory of sign<http://plato.stanford.edu/entries/peirce-semiotics/>

	Emoji	Eastern emoticon
Happy smile		
Why?!		
Massage head		




**Table 4: Some exclusive emojis and eastern emoticons**

In the following, we propose a typology of usages of emojis and emoticons. We first present different elements that may influence the usage of emojis, and the purpose underlying these usages.

#### Application context.

The choice to use emojis or emoticons may simply depend on the software context of display. Indeed, some limitations are inherent to the context. For instance in forums, emojis available can sometimes be seen as “ugly” and then emoticons will prevail. Also, a forum or any other platform could not have emojis implemented, leaving no choice to the user.

Table 5 demonstrates the impact the software context can have on an emoji’s usage. In this case, those pictures are referred under the same key: the “dancer” emoji. But the representations being too precise, the user could not use them in order to represent a dancer, and even for other purposes, as the Samsung and Twitter emojis convey opposite informations on the gender. A recent survey using 304 participants to complete 15 emoji interpretations demonstrate the same conclusion with an average of 37 interpretations per emoji<sup>12</sup> [16].



	Google	Twitter	Samsung
Dancer emoji			

**Table 5: dancer emoji: different representations based on software context**

Emojis and emoticons are widely used in Instant Messaging (IM), Text Messages (SMS) and chat rooms. We do not know yet if emojis are significantly present in email such as emoticons are, but in a study about 25% of emails contained emoticons [11].


**Formal versus informal context.** The emojis are mostly used in Informal Textual Conversations (called ITC) [11]. Few specific and standard emoticons are sometimes used in formal conversations (such as :-)). Indeed, the usage of emojis depends on the social relationship between the interlocutors, which is relevant to the external context.


**External context of the textual conversation.** An oral conversation may be carried on by the usage of an emoji. For instance the sentence “Yeah, the film was pretty good...” can be ambiguous without any additional information. How-

ever, “Yeah, the film was pretty good...” and “Yeah, the film was pretty good...” significantly reduce the ambiguity. In this context, or in another - a reply sentence “ofc...” for instance -, the usage of emojis can be tightly dependent to the previous oral conversation or to the event that occurred in the physical environment. Emojis are then given a function.

In these different contexts, we can identify the principal functions for which emojis are used:


**Sentiment expression.** A sentence can have a neutral emotional state. Therefore, using an emoji can simply add a emotion and a polarity to this sentence. An example of this could be “Ok, I’ll go there tomorrow.” compared to “Ok,

I’ll go tomorrow.”. The former only describe an action when the latter is more about complaining.

**Sentiment enhancement.** A sentence may convey different emotions. For instance, “I’m really happy and a little sad for her.”. The emojis may be used in this case to *enhance* the writer’s emotions. In other words, it allows the writer to disambiguate an emotional sentence by pointing at a specific emotion when several may be present. In the previous example, the addition of an emoji give the priority to one of the two emotions happiness and sadness: “I’m really happy and a little sad for her”.


**Sentiment modification.** Another objective attended by the writer when using emojis is to modify the emotions conveyed by the text or its intensity. Indeed, in Informal Text Communication (ITC), emojis can be used to express *complex* emotions such as sarcasm, irony, or non textual humor by simulating facial cues. To support this, in her study Kelly [12] came to the conclusion that emojis *surpassed* the text. With this, putting an emoji at the end of a text expressing the exact opposite emotion of the text allows the user to reproduce sarcasm, irony, etc.

It can also be used only to lightly modify the intensity of the conveyed emotion by lowering it. It is for instance the case when someone is complaining but with a humorous emoji

attached, such as “I’m so sad he is dead.”. In this case, the context will determine whether or not the emoji is used as an enhancer.

The usage of emojis for sentiment enhancement or sentiment modification is illustrated by the conclusions made by Novak *et al.* [15]: “on average, emojis are placed at the last tier of the length of a tweet”. Considering this fact, emojis are often used to enhance or modified a textual content by adding information at the emotional level, at the end of the sentence.

**Notifier.** Actually, emoji usages are not standardized at all and can be used as a simple notifier. As a way to keep the addressee’s attention the same way it is done in F2F communication: by body gestures through emojis. A good example of this is chat rooms in which people usually greet each other on first connection. Of course it can also be the case with SMS when interlocutors only want to keep in touch

by sending a, mostly, random emoji or emoticon ( o/ ,  , etc).

Thus, emojis can be used with a phatic function as one of the six functions Roman Jakobson described [10], and then have the sole objective to keep the conversation going. Kelly and Watts [13] showed user testimonial of this usage.

<sup>12</sup><http://grouplens.org/blog/investigating/-the-potential-for-miscommunication-using-emoji/>

**Convenience.** Emojis can be used to replace words and to give minimal information faster than by writing. Few research seems to have been conducted on this usage. The Emoji Sentiment Ranking corpus (ESR) [15] could be used to explore this usage. Indeed, this corpus indicates the average position of appearance of each emoji from 4% of 1.6 million tweets in 13 european languages. If a sentence is composed only of one emoji we can suppose that it is used to replace a word, used as a notifier, or to simply indicate a specific emotional state.

**Other: For fun.** Finally, another usage of emojis can be because the adresser think the emoji is fun and thus, that the addressee will find it fun too. But it is on the fringe of standard usages, if by standard we refer to sentiment related usages. This usage was also called “Permitted Play” by Kelly and Watts [13].

### 3.2 Subjective interpretation of emojis

In a linguistic point of view, an emoji is nothing more than a signifier in Saussure’s theory of a sign [5]. A signifier is the representation of the sign (e.g. the graphical image of the emoji), whereas the signified is the mental concept conveyed by the sign (e.g. the meaning of the emoji). As a signifier, an emoji can be interpreted differently by the addressee than the signified desired by the adresser. It is most likely that under specific circumstances the meaning of an emoji is different than under “normal” circumstances, following the theory of meaning from Palmer [19]. Palmer named it “context as meaning” when investigating on the meaning of a word by taking into account syntagmatic and paradigmatic contexts. This is why, as the main point of using emojis is to improve the message understanding, and to guide the addressee when different interpretations are possible, they enhance computer-mediated conversation (CMC) by providing contextual information [28] through facial cues and emotion signs. These signs are quite limited because their purpose is to reproduce F2F communication which also have limitations on its own: when facial expressions can be misinterpreted.

The interpretation of emojis is closely dependent of the first usage category we described before: the context of appearance. Coming back to the IM and SMS context, under normal circumstances the former is more synchronous than the latter because of the time spent by the message to arrive. When interpreting an emoji, users first consider the context of the interaction: application context, conversational context, social context etc. Thus, the context in which the emoji appears are necessary to truly understand and disambiguate it. Consequently, an emoji does not have a proper signification without it. Emojis do not have a universal understanding [12]. The previous example in Table 5 strengthen this idea.

## 4. EMOJIS IN SENTIMENT ANALYSIS

In this section we first tackle the available resources for emojis, and the other ones which could be relevant. And then we make an overview of the different methods relevant to emojis.

### 4.1 Available resources for emojis

Several resources exist for emoticons such as emoticon dictionaries<sup>13</sup> which goes from simple list of emoticons, to list

of polarity values (positive, negative, neutral) associated to emoticons. However, as far as we know there are only a few open-source resources for emojis: the Unicode list which contains 1624 emojis, and, the first emoji polarity lexicon (Emoji Sentiment Ranking<sup>14</sup> [15]) which is made of 751 emojis with their associated polarity annotated in context by human annotators.

Few emoticon text messages corpora such as the Mobile-ForensicTextMessageCorpus<sup>15</sup> for english, and the 88milms corpus<sup>16</sup> [20] for french, are available. However, except from Twitter, there is yet to be an emoji text corpus.

### 4.2 Emojis in Sentiment Analysis Models

In the following, we present a selected set of research works exploiting emoticons and emojis to improve sentiment analysis models. Indeed, sentiment analysis aims at developing systems to automatically recognize emotions, sentiments, or opinions in text. Given the context in which emojis or emoticons are used (Section 3), they represent particularly relevant cues for sentiment analysis since emoticons and emojis convey emotions.

As far as we know, emojis has not been used as often as emoticons in sentiment analysis studies. Emoticons, however, has been used a lot in association with Twitter based corpora, which represent the large majority of informal short text messages studies. Different types of approaches were used on emoticons: mainly keyword spotting and Machine Learning features. In a keyword spotting approach, emoticon dictionaries are used to extract the associated sentiment from emoticons. Then, in Machine Learning models, emoticons are considered as a feature for the model combined with other classical features such as hashtags in Twitter.

#### 4.2.1 Keyword spotting approach for emoticons

Keyword spotting is often applied with an external resource such as an emoticon dictionary or sentiment lexicon.

Related works combined emoticon lexicons and word sentiment lexicons to improve the supervised sentiment classification. For instance, Hogenboom *et al.* [9] proposed to detect the sentiment associated to emoticons by using synsets. These synsets regroup emoticons by sentiment such as happiness, sadness, etc. To do so, they combined a total of eight emoticon available lists such as Computer Users list<sup>17</sup>, The Canonical Smiley<sup>18</sup>, and The list of emoticons in MSN messenger<sup>19</sup>. If the emoticon was recognized in their emoticon sentiment lexicon, the emoticon’s sentiment override the words’ sentiments in the sentence.

Another example of keyword spotting was the one done by Neviarouskaya *et al.* [17]. They applied a rule-based method on chat rooms and used keyword spotting to take into account emoticons and abbreviations.

Kouloumpis *et al.* [14] applied keyword spotting by cre-

emoticons/ and <http://www.datagenetics.com/blog/october52012/index.html>

<sup>14</sup>[http://kt.ijs.si/data/Emoji\\_sentiment\\_ranking/](http://kt.ijs.si/data/Emoji_sentiment_ranking/)

<sup>15</sup><http://cybersecurity.cit.purduecal.edu/content/tmcorpus.html>

<sup>16</sup><http://88milms.huma-num.fr/>

<sup>17</sup><http://www.computeruser.com/resources/dictionary/emoticons.html>

<sup>18</sup><http://marshall.freeshell.org/smileys.html>

<sup>19</sup>[http://www.msgweb.nl/en/MSN\\_Images/Emoticon\\_list/](http://www.msgweb.nl/en/MSN_Images/Emoticon_list/)

<sup>13</sup>Some resources can be found here: <http://pc.net/>

ating “binary features”, i.e. in this case booleans, to distinguish positive, negative or neutral emoticons. This distinction was done along with intensifiers that they describe as all-caps and character repetition.

**Emoticon replacement.** More recently, emoticon lexicons were also used by Balahur [3]. They replaced the emoticon by their associated polarity, using the SentiStrength<sup>20</sup> emoticon sentiment lexicon. With this, they only retrieve whether an emoticon is negative or positive, and delete the neutral emoticons.

#### 4.2.2 Machine Learning approaches for emoticons

In machine learning approaches emoticons are used as a feature among others. The way this feature is used or represented differ from a research work to another.

**Emoticons as verified sentiment label.** Pak *et al.* [18] used a selected set of emoticons (mainly “:)” for positive, and “:(” for negative) to collect positive and negative sentiment tweets from Twitter. So they considered emoticons as verified sentiment label, i.e. label trustworthy enough for being used as a reference to train models. Meaning that, emoticon traduce sentence polarity. After using emoticons to lead the corpus creation, part of speech tags (PoS tags), i.e. verb, adjective, etc., were used as features to learn sentiment analysis models. They predicted PoS tags using the PennTreebank<sup>21</sup> tagset and the PoS tagger TreeTagger<sup>22</sup>. Then they examine the tag distributions by polarity in order to see if positive, negative, or neutral tweets have specific PoS tags patterns.

**Number of emoticons.** Becker *et al.* [4] used the numbers of positive, negative and neutral emoticons as features. They first used polarized bag-of-words features, representing a text as a set of words without order [7], each word being a feature. Then combined these features with the number of emoticons by polarity.

**N-grams features with emoticons.** In [26], Thelwall *et al.* used MySpace comments annotated by humans as a test set. Annotators were given a booklet containing emoticons with explanations. They determine the influence of emoticons on machine learning algorithms by comparing the SentiStrength<sup>23</sup> method with varying methods such as n-grams features, n-grams features with emoticons, etc. The more the data, the more emoticons created noise which decreased the sentiment analysis performance.

**Emoticons as binary features.** Binary features can be seen as booleans, and were used by Selmer *et al.* [24] after considering emoticons in the preprocessing phase. Then, with this binary representation of the text they used emoticon as another feature, another word.

Considering those approaches, emojis could be used to distinguish polarity of sentences just like emoticons. However, like emoticons, emojis can represent more than positive, negative, and neutral sentiments, but also fear, surprise, joy, etc. Even if Hogenboom [9] tackled these aspects for emoticons, it is yet to be done for emojis, which can convey more precise sentiments and emotions as we saw in Section 3.

However, in all those works, emoticons were rarely the

main point, but only taken as another additionnal feature or a noise to be taken care of.

#### 4.2.3 Approaches for emojis

As far as we know, no sentiment analysis model take into account emojis. Since emojis contains most of the emoticons. Some of these works could represent a good baseline for emojis as they allow to obtain premisses like the fact that Support Vector Machine [8] usually gives better results than Naive Bayes [23, 6] at least when using unigram features on temporal and domain dependency, but not on topic dependency. But the corpus used for these conclusions were different than the ones usually used for emojis (ITC). Moreover, emoticons were often reduced to a list of several western emoticons.

As for emojis, the emoticon based approaches for sentiment analysis could be repeated, but they ought to be used in a more specific way and with more granularity. Indeed, we previously presented the subtilities of the usages of emojis (Section 3), but those are slightly different from emoticon usages. Thus some emoticons caterogizations would not be sufficient. For instance, Hogenboom *et al.* [9] only took into account the “intensification”, “negation”, and “only sentiment” of an emoticon’s impact on sentiment value, which is more discrete that the one we presented in section 3. Plus, the n-gram approaches could not be relevant for emojis as they are only codes refering to images.

Related work on emoticons and emojis focused on the polarity of the text, considering the polarity of the emoticon, or considering the emoticon as a simple additionnal feature. We suggest that when this could have been really useful for sentiment polarity classification using emoticons, sentiment type identification using emojis would need to go further.

## 5. PERSPECTIVES

Most of the work exploiting emojis or emoticons for sentiment analysis consider only one usage of them: sentiment expression. In other words, they suppose that the emojis convey the emotions or sentiments of the entire sentence. However, emojis are also used for sentiment enhancement and sentiment modification. In sentiment analysis, sentiment enhancement and sentiment modification were never taken into account.

Given all the elements reported before (Sections 2, 3), we can assume emojis are very ambiguous and not reliable at all when taken out-of-the-box. Because they are ambiguous, emojis are rather difficult to use for sentiment analysis purposes. This is why it is important to identify the emoji usage type before using them. In the following we explore how emoji usage identification could be applied in sentiment analysis first. Then how emboddied conversational agents could benefit from emoji studies and recognition.

**Application in Sentiment Analysis.** In these perspectives, we first only consider three emoji usage cases: sentiment expression, sentiment enhancement, and sentiment modification. In order to identify those three usages, we make the premisses of a methodology using the first emoji sentiment lexicon (ESR [15]) and standard usages for sentiment analysis in short informal texts. Given a corpus made of informal text sentences containing at least one emoji, a methodology which goes by the following steps could be applied:

<sup>20</sup><http://sentistrength.wlv.ac.uk/>

<sup>21</sup><https://www.cis.upenn.edu/~treebank/>

<sup>22</sup><http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

<sup>23</sup><http://sentistrength.wlv.ac.uk/>

1. **Preprocessing:** by removing sentences only made of emojis then normalizing slang words and abbreviations using dictionaries<sup>24</sup>
2. **Sentence polarity detection:** Using an existing sentiment analysis system such as SentiStrength [27] which already ships several dedicated lexicons and resources. Also, we could use SVM with unigram features for words (bag-of-words) combined with sentiment lexicons such as SentiWordNet [2].
3. **Emoji polarity detection:** Using the ESR [15] to obtain the sentiment from emojis.
4. **Emoji polarity (3) and sentence polarity (2) comparison:** to distinguish sentiment expression (neutral to positive or negative), sentiment enhancement (emoji polarity equal to sentence polarity), or sentiment modification (emoji polarity different to sentence polarity).

The final step enables us to automatically label sentences with their usages: sentiment expression, enhancement, or modification. Then, this labeled corpus can be used to train a model to automatically identify the usages of a sentence.

Moreover, we can work on specific usage by considering a subset of the corpus containing only sentences and emojis used with a specific usage. This can be particularly useful in a sentiment analysis task. The sentiment modification emoji usage being contradictory by making the emoji polarity different from the sentence polarity, we can remove or ignore all sentences containing this label. Doing this we could significantly reduce the amount of ambiguous emoji usages, and lead to a more reliable train set. Fitting for training a classification model through SVM for instance.

**Application in human - ECA interaction.** Embodied Conversational Agents (ECA) or avatars may express emotions through their facial expressions. Emoji, often characterizing emotion through faces, may then be used to identify the agent's facial expressions in several ways:

- In a virtual environment, if the user types messages with emojis, the emojis representing faces can be used to automatically change the facial expressions of the avatar.
- During a chat dialog with an ECA, the emojis typed by the user can be exploited by the ECA as a cue on the emotion the user wants to express. Thus, emojis along to textual human content allows to access the facial expression the user wants to express. An ECA could be based on this information to map its showed expressions to the textual content the same way a human does. Plus, in real F2F conversations, human sometimes make mistakes or do not manage to show the desired facial expression. For instance when a person wants to show a smiling face without being in the mood to do it. With emojis this problem do not appear and makes them less complex.

We will then first develop the sentiment analysis approach and implement it in future work.

## 6. REFERENCES

- [1] A. Amalanathan and S. Margret Anuncia. Social network user's content personalization based on emoticons. 8(23).
- [2] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.
- [3] A. Balahur. OPTWIMA: Comparing knowledge-rich and knowledge-poor approaches for sentiment analysis in short informal texts. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM)*, volume 2, pages 460–465.
- [4] L. Becker, G. Erhart, D. Skiba, and V. Matula. Avaya: Sentiment analysis on twitter with self-training and polarity lexicon expansion. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM)*, volume 2, pages 333–340.
- [5] F. De Saussure. *Cours de linguistique générale*. Grande Bibliothèque Payot, critique edition.
- [6] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. 1:12.
- [7] H. Hamdan, F. Béchet, and P. Bellot. Experiments with DBpedia, WordNet and SentiWordNet as resources for sentiment analysis in micro-blogging. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM)*, volume 2, pages 455–459.
- [8] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf. Support vector machines. 13(4):18–28.
- [9] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak. Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 703–710. ACM.
- [10] R. Jakobson. *Essais de linguistique générale*.
- [11] T. A. Jibril and M. H. Abdullah. Relevance of emoticons in computer-mediated communication contexts: An overview. 9(4).
- [12] C. Kelly. Do you know what i mean >:( A linguistic study of the understanding of emoticons and emojis in text messages.
- [13] R. Kelly and L. Watts. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships.
- [14] E. Kouloumpis, T. Wilson, and J. D. Moore. Twitter sentiment analysis: The good the bad and the omg! 11:538–541.
- [15] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič. Sentiment of emojis. 10(12):e0144296.
- [16] H. Miller, J. Thebault-Spieker, S. Chang, I. Johnson, L. Terveen, and B. Hecht. “blissfully happy” or “ready to fight”: Varying interpretations of emoji.
- [17] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. EmoHeart: Conveying emotions in second life based on affect sensing from text. 2010:1–13.
- [18] A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 1320–1326.
- [19] R. F. Palmer. *Semantics*. Press Syndicate of the University of Cambridge, second edition edition.

<sup>24</sup><http://www.noslang.com/dictionary/>

- [20] R. Panckhurst. A large SMS corpus in french: From design and collation to anonymisation, transcoding and analysis. 95:96–104.
- [21] U. Pavalanathan and J. Eisenstein. Emoticons vs. emojis on twitter: A causal inference approach.
- [22] C. S. Peirce. Logic as semiotic: The theory of signs.
- [23] J. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop*, pages 43–48. Association for Computational Linguistics.
- [24] O. Selmer, M. Brevik, B. Gambäck, and L. Bungum. NTNU: Domain semi-independent short message sentiment classification. page 430.
- [25] E. R. Team. Emoji\_report.
- [26] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. 61(12):2544–2558.
- [27] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. 61(12):2544–2558.
- [28] C. C. Tossell, P. Kortum, C. Shepard, L. H. Barg-Walkow, A. Rahmati, and L. Zhong. A longitudinal study of emoticon use in text messaging from smartphones. 28(2):659–663.