

Detecting Facial Emotions to Support Linguistic Analysis of Political Tweets: a Multimodal Approach

Julien Longhi

▶ **To cite this version:**

Julien Longhi. Detecting Facial Emotions to Support Linguistic Analysis of Political Tweets: a Multimodal Approach. 8th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC-Corpora 2021), Oct 2021, Nijmegen, Netherlands. hal-03405088

HAL Id: hal-03405088
https://hal.science/hal-03405088

Submitted on 27 Oct 2021

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.

Proceedings

of the 8th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora (CMC-Corpora 2021)

28-29 October 2021

Radboud University, Nijmegen

Proceedings of the 8th Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora 2021)

28-29 October 2021, Radboud University (Nijmegen, the Netherlands)

Editors: Iris Hendrickx, Lieke Verheijen, and Lidwien van de Wijngaert

Published by: Radboud University
Nijmegen, 2021

Conference website: https://applejack.science.ru.nl/cmc2021/

This work is licensed under a Creative Commons “Attribution 4.0 International” license.

Detecting Facial Emotions to Support Linguistic Analysis of Political Tweets: a Multimodal Approach

Julien Longhi
CY Cergy Paris université, AGORA, IDHN, IUF
E-mail: julien.longhi@cyu.fr

Abstract

This paper focuses on the analysis of audiovisual content embedded in tweets, especially in a political context. It is based on research conducted since 2013 on digital discourse on Twitter, and addresses a new challenge for the CMC community: the automatic consideration of multimedia content embedded in messages. To this end, we propose a promising research direction: the automatic extraction of emotions from political personalities, using artificial intelligence. Although this work is still in its early stages, it offers promising possibilities for discourse analysis, while requiring some precautions. It also allows us to consider new solutions for the linguistic analysis of emotions or feelings, by providing additional labels for annotating corpora.

Keywords: political tweet, sentiment analysis, emotions, deep learning

1. Introduction

Social media are increasingly integrating multimedia content. This is particularly the case for tweets which in their body contain, in addition to text, visual or audiovisual elements (Simon et al., 2016) that contribute to the argumentation put forward. For example, in the following tweet, an extract from a speech by presidential candidate Emmanuel Macron is quoted by the user, but the quote is linked to its broader context in the form of a video:



Figure 1: Capture of a tweet integrating a video of Emmanuel Macron

In this paper, we extend our work conducted since 2013 on digital discourse on Twitter (Longhi 2013; Djemili et al. 2014; Longhi 2020) and address a new challenge for CMC corpora research: the automated consideration of multimedia content embedded in social media messages. To this end, we propose a promising research area whose scientific progress is remarkable: the automatic extraction of emotions from political figures, based on tools using artificial intelligence. This exploratory paper is structured as follows: first, we delimit the main lines of the research by situating ourselves in relation to different approaches; second, we briefly present some criteria for describing faces, using an example from Emmanuel Macron's video; third, we look at their use for identifying emotions and propose a multimodal analysis of a video sequence by bringing together several observations; fourthly, we add to these results a linguistic analysis based on BERT, a deeply bidirectional, unsupervised language representation. We then conclude on the significance of such a study (which extracts meaningful elements from videos that are useful for linguistic analysis) and present a way to encode emotions in corpora using the TEI format.

2. Delimitation of the study area

According to Scherer et al. (2011: 426), "emotion expression and impression are determined by an interaction of psychobiological and sociocultural factors", and "the question of whether [the] expression and impression of emotion are either universal or culturally specific in terms of a dichotomy is therefore moot, as science has progressed to a higher level of understanding". Of course, cross-cultural issues related to the automatic detection of emotion need to be addressed, and ontological positions related to the labelling of multimodal and linguistic content from the outside and from the participants' point of view which we will integrate also need to be explored.

To do so, we plan to develop the etic and emic perspectives in investigating how emotions are expressed and interpreted by others. Indeed, the emic/etic concepts, coined by Pike, are "two basic standpoints from which a human observer can describe human behavior, each of them valuable for certain specific purposes" (Pike, 1954: 8). Mostowlansky & Rota (2020) point out that, according to Pike, "an etic approach would rely on a generalised classification system devised by the researcher in advance for the study of any particular culture in order to compare and classify behavioural data from across the world", while "an emic approach would dispense with a priori means of classification". For the moment, our research has been focusing on one culture at a time, but for future work we intend to integrate an emic approach (which would "call attention to the fact that two etically identical behaviours can in fact differ profoundly, depending on the meaning and purpose of the actors").

3. Recognition of action units

Soysal et al. (2017) propose "a new deep learning scheme that integrates convolutional neural networks with association rule mining", according to the "action units" (AU) described on faces:

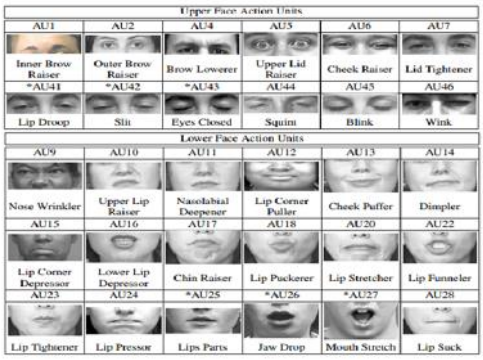


Figure 2: Action units (Saysal et al., 2017)

We can see above an inventory of different facial expressions, for both the upper and lower face, which we can apply to our video as below:



Figure 3: Eyebrow (left), Mouth (right)

4. Video analysis

To do an automatic analysis of this video, we used OpenFace[1] which provides "free and open source face recognition with deep neural networks". We applied this tool to a much commented "punchline", *Madame Le Pen, la France mérite mieux que vous* ("Madame Le Pen, France deserves better than you"), lasting 3 seconds. When analysing the video, OpenFace (Amos et al., 2016) proposes a face model with a large number of points:

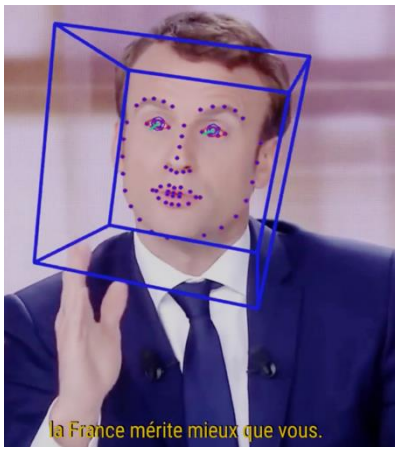


Figure 4: Face modelling

The video is therefore decomposed into a set of images, and each is automatically analysed according to the AU described in Figure 2.

4.1. Segmentation and interpretation of results

Once this division is made, we can observe the different AUs on graphs:



Figure 5: Presence of AU according to the timestamp of a sequence

Thus, we can observe the "strength" of each AU by "frame", which then allows us to make assumptions about potential facial emotions. Indeed, as some work shows, the combination of different AUs potentially allows us to define some basic emotions, as illustrated for example by Wegrzyn et al. (2017):

[1] https://cmusatyalab.github.io/openface/. AU extraction with OpenFace was performed with Aymeric Erades, research engineer at IDHN in 2021.

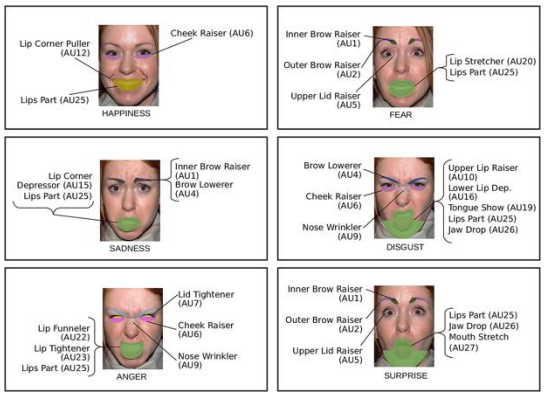


Figure 6: Definition of six emotions by Wegrzyn et al. (2017)

The hypothesis we make is that we could detect emotions in order to add this information to the analysis of the messages posted. To do so, and in an ecological approach to corpora, it is necessary to be able to return to the image corpus and verify the automatic results.

4.2. Methodological precaution: categories of emotions and using the corpus

According to Crawford (2021), there are several limitations to emotion detection:

- The emotion categories themselves are problematic: they are often based on Ekman's pioneering work and "we should question the origins of those categories themselves, as well as their social and political consequences". For us, therefore, this work of detecting emotions comes in "support" of linguistic analysis and is not an end in itself;
- As far as detection is concerned, "Emotions are complicated, and they develop and change in relation to our cultures and histories": it is therefore important to be able to return to the corpus to observe the data and validate the results.

For this purpose, OpenFace allows us to return to the identified frames and related images:



Figure 7: Back to the video corpus

We can thus prevent potential mistakes being made by the system and observe precisely some key moments whose interpretation would seem to be relevant.

5. For a multimodal analysis of political tweets

This work allows us to partially overcome the difficulty related to the multimodal dimension of the messages written on Twitter. Indeed, in the case of the quote that interests us here, Macron's account limits its sharing to the textual content only:



Figure 8: Tweet of Emmanuel Macron

This message is interesting from a linguistic point of view: to quantify it, we applied two models developed with BERT to capture the intensity or quality of the sentiments expressed. We chose BERT because, according to Tenney et al. (2019), "deep language models can represent the types of syntactic and semantic abstractions traditionally believed necessary for language processing, and moreover that they can model complex interactions between different levels of hierarchical information". In particular, "BERT is the first deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus".[2] Even if sentiment analysis is a burning issue, Essebar et al. (2021) demonstrate that "French PTM shows improvements of text representation in many NLP tasks including sentiment analysis at sentence-level".

We thus applied two pre-trained models based on BERT and when we looked at our sentence we found that it rather expressed anger, with strong intensity:

[2] https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

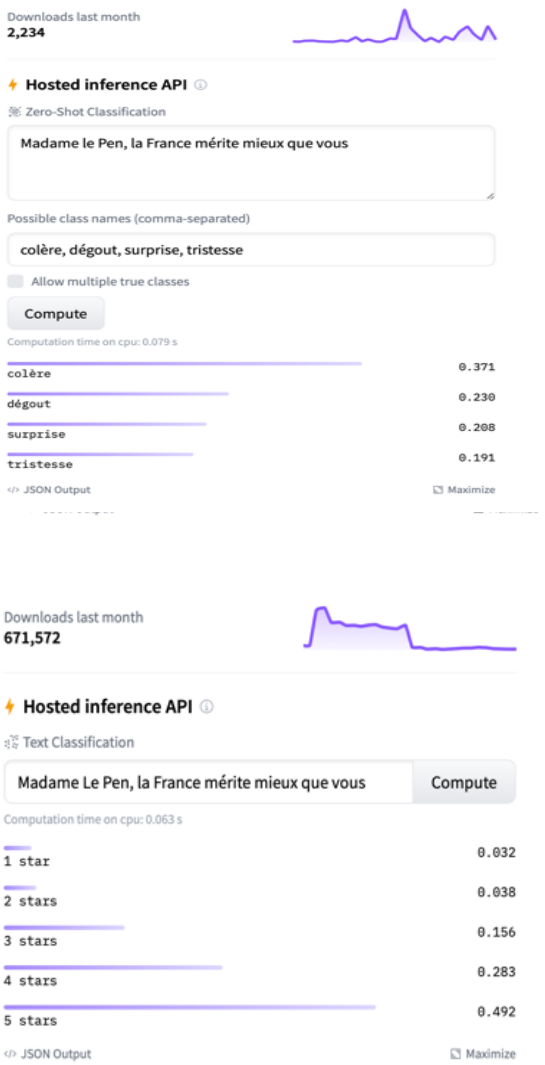


Figure 9: Linguistic analyses based on BERT models

The sentence is identified as anger (with a score of 0.371) with an intensity of "5 stars" (0.492). But since, as we have seen, this statement was part of an interaction, we can try to know if the candidate also physically expressed this anger. In figure 5, we proposed a graph, called AU50_r, which represents the detection of anger (by combining AUs 6, 7, 9, 22, 23, 25):

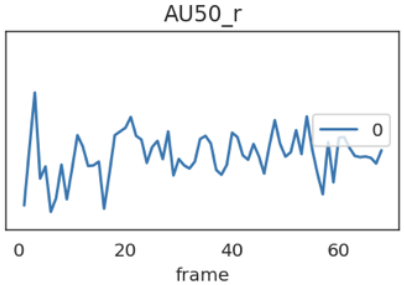


Figure 10: Representation of the emotion "anger" in the short extract

We can now confirm that the automatic analysis of the candidate's face reflected a certain anger. This result met our initial objective of introducing a multimodal dimension to the automated analysis of political messages on Twitter. In addition to enriching the corpus, this solution for the linguistic analysis of emotions or feelings (by providing additional labels for annotating corpora) could be a precise help for automatic language processing by making the annotation of corpora reliable, which could then be used to train models with the aim of improving emotion or feeling recognition algorithms.

6. Conclusion

This paper, which opens up new perspectives of research for the linguistic analysis of political tweets, allows us to obtain information from the multimedia content embedded in tweets, in order to go beyond the limitations of a sole textual analysis. We can imagine, in the short term, that these analyses could be integrated as annotations in corpora, with a set of basic emotion labels or general descriptors. We could thus have, in the same way as temporal or material variables, an "emotion" variable using an xml tag, for example, in the TEI representation of tweets:

<post xml:id="cmr-présidentielle-a859882207648854016" who="#cmr-présidentielle-p1976143068"
when="2017-05-03T23:27:48.0" xml:lang="fra">

Madame Le Pen, la France mérite mieux que vous. <distinct type="twitter-hashtag"><ident>#</ident><rs ref="https://twitter.com/search?q=%232017LeDébat&src=hash">2017LeDébat</rs></distinct>

<trailer>
<fs>
<f name="medium">
<string>Twitter Web Client</string>
</f>
<f name="favoritecount">
<numeric value="9828"/>
</f>
<f name="retweetcount">
<numeric value="9625"/>
</f>
<f name="emotion">
<emotion name="anger"/>
</f>
</fs>
</trailer>

This would allow, at first, to take into account the facial emotion expressed when such or such a sentence is uttered. Of course, this opens the way to many other aspects (gestures, mimics) on which we will work in future research.

More generally as far as the analysis of political discourse is concerned, this perspective allows for a more complete grasp of political argumentation since, according to Lara et al. (2016: 155), "emotion is not only a dimension working

as an enhancer or mitigator for the persuasive aim", "it is used as an argument itself". This allows us to better understand political discourse, especially on Twitter, by not neglecting an important part of the contribution made by the staging of emotions. To improve that, future work should include intercultural specificities because, "since emotion recognition is culturally dependent, it is predicted that a culturally inclusive recognition model should yield better results than a model that does not consider culture" (Quiros-Ramirez & Onisawa, 2015). This is another challenge to be taken up in order to offer a better match between the analysis model and possible application contexts.

7. References

Amos, B., Ludwiczuk, B., Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications. CMU-CS-16-118, CMU School of Computer Science, Tech. Rep.

Crawford, K. (2021). Artificial intelligence is misreading human emotion, The Atlantic, https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/

Djemili, S., Longhi, J., Marinica, C., Kotzinos, D., Sarfati, G.E. (2014). What does Twitter have to say about ideology?. In *NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication/Social Media-Pre-conference workshop at Konvens 2014*, Universitätsverlag Hildesheim.

Ekman, P. (1984). Expression and the nature of emotion. *Approaches to emotion*, pp. 319-344.

Essebbar, A., Kane, B., Guinaudeau, O., Chiesa, V., Quénel, I., Chau, S. (2021). Aspect Based Sentiment Analysis using French Pre-Trained Models. In *ICAART* (1), pp. 519-525.

Lara, E.A., Márquez, A.C., Fuentes-Rodríguez, C. (2016). Emotional argumentation in political discourse. *A gender-based approach to parliamentary discourse*, 68, pp. 129-159.

Longhi, J. (2013). Essai de caractérisation du tweet politique. *L'information grammaticale*, 136, pp. 25-32.

Longhi, J. (2020). Proposals for a discourse analysis practice integrated into digital humanities: theoretical issues, practical applications, and methodological consequences. *Languages*, 5(1), 5.

Ma, J., Li, X., Ren, Y., Yang, R., Zhao, Q. (2021). Landmark-Based Facial Feature Construction and Action Unit Intensity Prediction. *Mathematical Problems in Engineering*, https://www.hindawi.com/journals/mpe/2021/6623239

Mostowlansky, T., Rota, A. (2020). Emic and etic. *Cambridge Encyclopedia of Anthropology*, https://www.anthroencyclopedia.com/entry/emic-and-etic

Pike, K.L. (1954). *Language in relation to a unified theory of the structure of human behavior, part 1*. Glendale, Calif.: Summer Institute of Linguistics.

Quiros-Ramirez, M.A., Onisawa, T. (2015). Considering cross-cultural context in the automatic recognition of emotions. *International journal of machine learning and cybernetics*, 6(1), pp. 119-127.

Scherer, K.R., Clark-Polner, E., Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology*, 46(6), pp. 401-435. https://doi.org/10.1080/00207594.2011.626049

Simon, J., Bénédicte, T. (2016). Le poids argumentatif des tweets multimodaux dans les récits médiatique. In *Colloque international Médias numériques et communication électronique*, Klog, pp. 227-238.

Soysal, Ö. M., Shirzad, S., Sekeroglu, K. (2017). Facial action unit recognition using data mining integrated deep learning. In *2017 International Conference on Computational Science and Computational Intelligence* (CSCI), pp. 437-443.

Tenney, I., Das, D., Pavlick, E. (2019). BERT rediscovers the classical NLP pipeline. arXiv preprint arXiv:1905.05950.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS one*, 12(5).

Models:

https://huggingface.co/nlptown/bert-base-multilingual-uncased-sentiment?text=Madame+Le+Pen%2C+la+France+mérite+mieux+que+vous

https://huggingface.co/BaptisteDoyen/camembert-base-xlni?candidateLabels=colère%2C+dégout%2C+surprise%2C+tristesse&multiClass=false&text=Madame+le+Pen%2C+la+France+mérite+mieux+que+vous