Master's thesis

A sentiment analysis of Spanish and Italian news articles about COVID-19

Exploring the emotional reaction on government-applied restrictions

A thesis submitted in fulfilment of the requirements for the degree of

Master of Arts

by

Jordany Werzner Regalado

First Examiner: Prof. Dr. Michael Paul Robert Richter
Second Examiner: PD Dr. Anna Ladilova
Language, Literature, Culture, M.A. (4th semester)
Department of German Studies
Justus Liebig University Giessen
Date and place of submission: Giessen, 20.11.2021

Abstract

In 2020 the effects of the COVID-19 pandemic and the related government-applied nationwide measures deeply influenced the Italian and Spanish population, not only financially and socially, but also in terms of the emotional state of all individuals. Due to social isolation, social media and other online communication platforms served as an outlet to express opinions and feelings about the situation and to interact with other peers. This resulted in the availability of a great amount of online data that can be utilised to analyse the sentiment and emotions expressed during this time. The aim of this study was to use this data to perform a context-based sentiment and emotion analysis in order to determine a relationship between shifting changes of sentiment and emotion and the application restriction measures in Italy and Spain. Most current studies focus on English data extracted from Twitter and show little interest towards other data sources and other languages. This thesis therefore based its analysis on over 700 000 Spanish and Italian comments extracted from the newspapers *La Repubblica* and *El Mundo* using a Python-based web crawler. The created press corpora represent another user demographic and can be used for the analysis of the Italian and Spanish language in other research areas. A sentiment classification algorithm based on one-layer convolutional network was used to determine the polarity in the comments. The classification method achieved an F-Score of 0.87 for the Spanish language model and an F-Score of 0.81 for the Italian model. The emotion detection was performed using the Syuzhet R Package and NRC Emotion Lexicon to create emotion scores during different time frames. Using a graphical analysis the study determined an existing emotional reaction, that could be put in relation to the measures applied by the governments. Furthermore, utilising Pearson's correlation coefficient, it was determined that Robert Plutchik's theory of emotional opposites applies to the analysed context. Moreover, a negative correlation was detected between the level of trust and the level of fear expressed in comments mentioning the government. All results were shown in a contrastive manner to compare the emotional reactions in the comment sections.

The insights gained from this study can be used for cultural, linguistic and political analysis, political decision-making and for the development of strategies to manage the pandemic and other national catastrophes, while considering the emotional state of the population.

Abstract

Im Jahr 2020 hatten die COVID-19-Pandemie und die damit zusammenhängenden landesweiten Maßnahmen der Regierung tiefgreifende Auswirkungen auf die italienische und spanische Bevölkerung, nicht nur in finanzieller und sozialer Hinsicht, sondern auch in Bezug auf den emotionalen Zustand jedes Einzelnen. Aufgrund der sozialen Isolation dienten die sozialen Medien und andere Online-Kommunikationsplattformen als Ventil, um Meinungen und Gefühle über die Situation zum Ausdruck zu bringen und sich mit anderen Individuen auszutauschen. Dadurch steht eine große Menge an Online-Daten zur Verfügung, die zur Analyse der in dieser Zeit geäußerten Gefühle und Emotionen genutzt werden können. Ziel dieser Studie war es, diese Daten zu nutzen, um eine kontextbasierte Stimmungs- und Gefühlsanalyse durchzuführen, um eine Beziehung zwischen den Veränderungen der Stimmung und Emotionen und den Regierungsmaßnahmen in Italien und Spanien zu ermitteln. Die meisten aktuellen Studien konzentrieren sich auf englische Daten, die aus Twitter extrahiert wurden, und zeigen wenig Interesse an anderen Datenquellen und anderen Sprachen. Diese Arbeit stützt sich daher auf über 700 000 spanische und italienische Kommentare, die mit Hilfe eines Python-basierten Web-Crawlers aus den Zeitungen *La Repubblica* und *El Mundo* extrahiert wurden. Die erstellten Pressekorpora repräsentieren eine andere Nutzerdemografie und können für die Analyse der italienischen und spanischen Sprache in anderen Forschungsbereichen verwendet werden. Zur Bestimmung der Polarität in den Kommentaren wurde ein auf einem einschichtigen Faltungsnetzwerk basierender Stimmungsanalyse-Algorithmus verwendet. Die Klassifizierungsmethode erzielte einen F-Score von 0,87 für das spanische Sprachmodell und einen F-Score von 0,81 für das italienische Modell. Die Erkennung von Emotionen wurde mit dem R-Paket Syuzhet und dem NRC Emotion Lexicon durchgeführt, um Emotions-Scores für verschiedene Zeiträume zu erstellen. Mithilfe einer grafischen Analyse wurde in der Studie eine bestehende emotionale Reaktion ermittelt, die mit den von den Regierungen angewandten Maßnahmen in Beziehung gesetzt werden konnte. Darüber hinaus wurde mit Hilfe des Pearson-Korrelationskoeffizienten festgestellt, dass Robert Plutchiks Theorie der emotionalen Gegensätze auf den untersuchten Kontext zutrifft. Außerdem wurde eine negative Korrelation zwischen dem Grad des Vertrauens und dem Grad der Angst festgestellt, die in den Kommentaren über die Regierung zum Ausdruck kommt. Alle Ergebnisse wurden

kontrastiv dargestellt, um die emotionalen Reaktionen in den Kommentarbereichen zu vergleichen.

Die aus dieser Studie gewonnenen Erkenntnisse können für die kulturelle, sprachliche und politische Analyse, für die politische Entscheidungsfindung und für die Entwicklung von Strategien zur Bewältigung der Pandemie und anderer nationaler Katastrophen unter Berücksichtigung des emotionalen Zustands der Bevölkerung genutzt werden.

Preface

I submitted this thesis for the finalisation of my master's programme in language, literature and culture with computational linguistics as a major subject in the department of German studies at JLU. It deals with the context-based sentiment and emotion analysis of the comments of COVID-19 articles in the newspapers *EL Mundo* and *La Repubblica* and aims to determine whether there is a relationship between the shifting sentiment and measures applied by the Italian and Spanish government.

All of us had to react to a worldwide emergency in a short span of time and used online communication as an outlet to express our emotions. Considering the situation, dominating emotions were, of course, negative. The sadness, fear and the anger caused by the pandemic describe only a small spectrum of all emotions we felt. All the more, we should also remember the positive moments of joy, our trust in the applied measures and the messages of hope sent to all of us through social media. Posts about the pandemic had rapidly covered a broad range of emotions. This gave me the idea of analysing the available data. Since most researchers analyse English data, I came across difficulties when trying to find adequate datasets for sentiment analysis and for the training of Machine and Deep Learning algorithms in the target languages. This shows that sentiment and emotion analysis research is still at an early stage and more data is needed, especially for languages other than English. I hope that this thesis can contribute to fill this gap and provide another perspective by integrating cultural and political insights.

I want to express my gratitude towards my first examiner Prof. Dr. Michael Paul Robert Richter and my second examiner PD Dr. Anna Ladilova for helping overcome all obstacles and guiding me through the writing process with their scientific knowledge. Furthermore, I wish to acknowledge the support provided by my former lecturer Mr. Holger Grumt Suárez during the creation of the press corpora. I would also like to especially thank my fellow student Daniel Baldassare and my brother Eliezer Werzner Regalado for their unconditional assistance and guidance throughout my studies.

Finally, I would like to dedicate my thesis to my parents, my brothers, my sister and my wife Franziska Werzner Regalado for their continuous support and encouragement.

Giessen, 20.11.2021 Jordany Werzner Regalado

Table of Contents

1 Introduction

The ongoing COVID-19 pandemic has caused an international public health crisis, that will affect the world's economy, culture, and society in the long term. In 2020, government institutes and entities were challenged with providing a suitable solution to stop the spread of the coronavirus, aiming to ease the burden on the public health system and thus save the lives of those that were and could be dangerously infected. The first two European countries to report a serious rise of infections and deaths were Italy and Spain. These countries quickly became the focal point for news reports in the European and worldwide press. On the one hand, during the first nationwide lockdown, images of overwhelmed hospital staff in Italy dominated world news with the country becoming the epicentre for the virus. On the other hand, Italian citizens also provided images of hope keeping up the morale, by singing from their balconies and sending messages of hope through social media. At the same time, concern rose in the Spanish government and population as the situation in Italy worsened and the virus gained prominence in their own territory. In a short time span, the virus changed the financial and social situation of most individuals, while also affecting their emotional well-being.

Due to social isolation, the population resorted to online communication to interact with other people and used social media platforms, such as Twitter, to express their feelings and opinions about the situation. These platforms thus build the basis for many research papers, which aim to exploit the available data to analyse the expressed sentiment and emotion using sentiment analysis and emotion detection. Sentiment analysis means the automated extraction of feelings and opinions expressed in text and multimodal data, while emotion detection focuses on the detection of emotion expressed in a document.

The field of sentiment analysis and emotion detection is a rather new research field., which is gaining more and more attention in research areas beyond information technology. Apart from its application in commercial business, it holds great potential for the analysis of sociocultural phenomena and for the support of political decisions, which need to consider the emotional state of the population.

Most of the studies released on sentiment analysis about the pandemic use Twitter data written in the English language and pay little attention and interest towards other languages and the analysis of comment sections in online newspapers. Compared to popular platforms, such as Facebook or Twitter, these comment sections are less susceptible to the spread of false information and still provide a great source of textual and multimodal data. Users of online news sites represent another demographic, which not only uses the platforms as a reliable source of information, but also have the possibility to comment on the articles directly after the reading process.

This study aims to use the available data to determine whether sentiment or emotion shifts during pandemic were related to the restriction measures applied by the Italian and Spanish government. After providing a thorough summary of the applied measures, a Python-based crawler was built to gather textual data about the pandemic from the newspapers *La Repubblica* and *El Mundo* and thus build two corpora. These corpora will be referred to as "press corpora". The gathered information will be used to track the emotions and feelings expressed by the readers. Therefore, a sentiment analysis will be performed to determine the negative or positive sentiment expressed in the comment section. The sentiment polarity will be classified using a classification algorithm based on a single-layer convolutional neural network using the Python library spaCy. The algorithm will use pre-labelled training data obtained from Twitter for the classification of untrained comments from the press corpora. The results will be put into the context of the government applied measures to determine whether they caused sentiment shifts. This will be done by using a graphical analysis, which includes the time sections of the applied measures and the sentiment found during these time frames. A special focus will be laid on the first nationwide lockdown applied by both countries. The sentiment analysis will be followed by an emotion analysis. The emotion detection will be performed using the Syuzhet R package and integrating the NRC Lexicon to obtain emotion scores for different time frames. The shift of emotions will be put into the context of government applied measures by using a graphical analysis as well. Hereafter, this study will determine whether there is an existing relationship between emotional opposites, as proposed by Robert Plutchik. To answer this question the correlation of the scores of emotional opposites will be calculated to determine if there is a negative linear relationship. In addition, the last chapter will be dedicated to the analysis of trust in comments mentioning the government and government entities. All results will be shown in a contrastive manner to determine whether the emotional reactions in both countries were similar or differed from each other.

2 Current state of research

In the period from January 2020 to December 2020 the publications of research papers dedicated to sentiment analysis on the COVID-19 pandemic were limited due to the lack of available data. Most research papers therefore used the Twitter API, since it could deliver instant information for the analysis of the public sentiment. The number of papers and datasets is however steadily growing, with the focus generally switching to more concrete topics, such as the effects of the pandemic on the economy, the spread of misinformation or the sentiments and feelings towards vaccines.

Twitter Sentiment Analysis on Coronavirus using Textblob was one of the first studies on the subject and was published by Kaur and Sharma on March 16, 2020. Using the libraries Tweepy[1] and Textblob[2] they performed a sentiment analysis on 2058 Tweets about the pandemic. Their analysis delivered a high number of neutral Tweets with a total of 43.9 % (cf. Kaur et al. 2020, p. 9-12). On the work *COVID-19 pandemic: a sentiment analysis: A short review of the emotional effects produced by social media posts during this global crisis* from July 17, 2020, Amish Kumar et al. gave a first impression on the sentiments and emotions found on social media after the first major wave of the pandemic. They used the Twitter API to analyse over a million English tweets and performed a sentiment analysis and emotion detection using a mixture of lexicon-based and statistical approaches. They determined an overall positive sentiment and a high amount of trust, while fear was the highest negative sentiment (cf. Kumar et al 2020, p. 2). Also published in July 2020, the paper *Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic* by Kruspe et al. focused on the analysis of the sentiment of tweets using a combination of BERT[3] and word and sentence embedding. They determined that the general sentiment was very negative at the beginning of the pandemic and became more positive over time. The negative sentiment was generally higher than the average sentiment (cf. Kruspe et al. 2020, p. 8-10). Michele Costola et al. analysed the impact of COVID 19 news on the stock market in the paper *Machine learning sentiment analysis, COVID-19 news and stock market reactions* published in September

[1] Tweepy is a python library used to access the Twitter API.
[2] TextBlob is a python library mainly used for the analysis of textual data.
[3] BERT stands for Bidirectional Encoder Representations from Transformers and is language model which can be used to create state-of-the-art models for a variety of language tasks (cf. Devlin et al. 2018, 1).

2020. They collected different online articles and performed a sentiment analysis using the BERT model. Using this approach, they determined a relationship between the determined sentiment and the return on the financial market (cf. Costola et al. 2020, p. 9-15). Furthermore, Vijay et al. analysed the polarity of Twitter Data using Textblob and NLTK[4] in their paper *Sentiment Analysis on COVID-19 Twitter Data*, which was published on IEE Explore in February 2021. The study determined a general positivity in Indian Tweets regarding both the pandemic and the measures applied by the Indian government (cf. Vijay et al. 2021, p. 5-7). Analysing Twitter Data as well, Singh et al. published the article *Sentiment analysis on the impact of coronavirus in social life* on March 19, 2021, to further analyse the social sentiments expressed in social media during the pandemic in India. They used Vader Sentiment Analyser[5] for the determination of the polarity, while also considering the quantity of retweets and likes and used the BERT model for emotion classification. While the general sentiment on Twitter was positive in general, the emotion detection also served to determine the success or failure of certain measures applied by the Indian government.

Other works focused their studies on the detection of emotion. In the paper *Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions* from June 2020, Li et. al. use Twitter and Weibo to determine the emotions expressed in these platforms during the pandemic using the description-based BERT model (cf. Li et al. 2020, 4).

Published in July 2020 by Aslam et al., the study *Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak* classified the sentiment and emotion expressed in news headlines obtained from English news sites using the R package “sentiment” and NRC Lexicon (Aslam et al. 2020, 4). The study determined a highly negative sentiment, while fear and trust were the dominant emotions detected.

Das et. al also published the study *Characterizing public emotions and sentiments in COVID-19 environment: A case study of India* in July 2020, which is based on the analysis of Indian Tweets. The sentiment was classified using sentiment-based topic models, while the emotion detection was performed using the NRC Lexicon as well. Imran et al. showed a sentiment and emotion analysis from a cultural perspective in their

[4] The Natural Language Toolkit (NLTK is a suite of open-source programme modules, tutorials and problem sets, providing ready-to-use computational linguistics courseware (Bird et al. 2014, 1).

[5] VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains (Hutto et al. 2014, 1).

work *Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets* by using deep long short-term models and integrating emoticons to analyse Tweets. The study showed that Tweets from USA, Canada, Pakistan were highly correlated, while opposite trends were detected in Sweden in Norway (cf. Imran et. al. 181085 f.).

Except for Costola et al. and Aslam et. al., most of the works used Twitter data to determine the sentiment of the public on different subjects.[6] This is facilitated by the easy use of the Twitter API and the large amount of Twitter datasets, that can be accessed and downloaded online. Nevertheless, most Twitter sentiment analysis works don't consider the limits of Twitter data with regards to the representation of the public. A study of Pew Research Center determined that Twitter users in the USA are generally younger, wealthier and tend to have a more liberal political view (cf. Wojcik et. al. 2019, n.p.). This means that the used data is skewed towards a specific demographic. Moreover, Tweets contain a high number of fake tweets and false information, spread by automated bots, which can be difficult to exclude from the dataset. In addition to that, Twitter users are obligated to limit their tweets to a certain number of characters, which limits the expression of the train of thought. This limitation also results in a more frequent use of abbreviations and word omissions.

Considering the limits of Twitter data, this study aims to provide a dataset with data obtained from the comments of the news sites El Mundo and La Repubblica. These newspapers were chosen to cover a more central political spectrum, with El Mundo generally leaning towards the centre-right, while La Repubblica generally has a centre-left stance. The comment sections of these news sites don't contain any character limit and are less susceptible to the automated spread of false information, while still providing useful information for linguistic analysis. Furthermore, since most studies perform sentiment analysis on English texts, the extracted will contribute to research in other languages. The sentiment analysis and emotion detection performed in this study will therefore have a different focus group, namely the readers of official news sites and will focus on the Spanish and Italian Language. The data will be used to track the direct reaction of the readers shortly after having read the changes of the state of the pandemic and the official measures applied by the Italian and Spanish government in the year 2020.

[6] Many works are still being released at the time of writing and may include other data sources.

3 The pandemic in Italy and Spain in 2020

As shown in figure 1, in 2020 the development of the pandemic in Italy and Spain was similar.

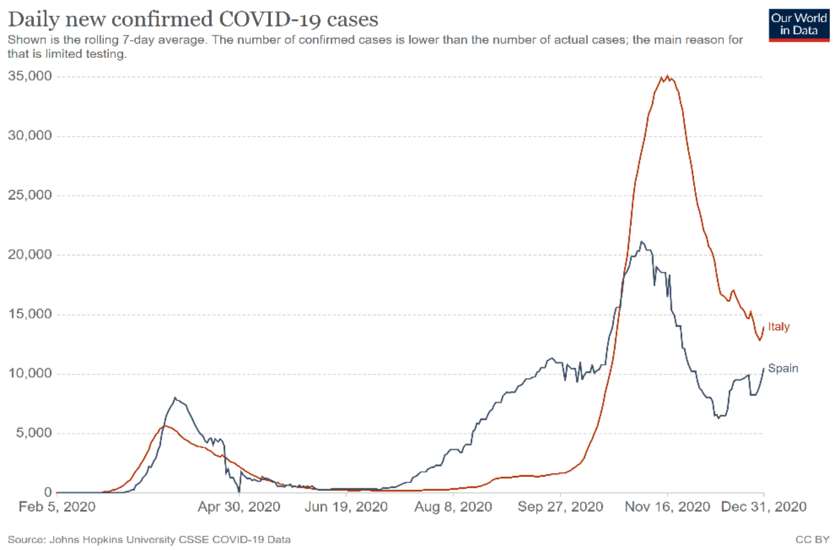

Figure 1 - COVID-19 in Spain and Italy 2020[7]

Italy had its first case on January 23, 2020, approximately a month after the first confirmed case in China. It was reported that two Chinese tourists had brought the virus to Italy (cf. Severgnini 2020, n.p.). On January 30, after the first case was detected, the WHO classified the pandemic as a Public Health Emergency of International Concern (PHEIC) and proposed recommendations to stop the spread of the virus after the first cluster of cases was confirmed in Wuhan, China (cf. WHO 2020, n.p.). The Italian government reacted quickly by declaring the state of emergency on the same day and suspended all flights from and to China (cf. Sclaunich 2020, n.p.). Within the same period the first infected person was detected in the city of La Gomera in Spain, after a German tourist had tested positive. Nevertheless, the Spanish government decided against the declaration of the state of emergency and to continue adopting their own protocol, which mainly consisted of isolating infected individuals (cf. Linde 2020, n.p.).

[7] Retrieved from: https://ourworldindata.org/coronavirus [Online Resource]

3.1 National containment measures during the first wave

By February 21, 2020, the first cluster of 16 cases was reported in Lombardy and the first death was reported in the region of Veneto. The Italian government put 10 towns in northern Italy under lockdown, affecting about fifty thousand people (cf. La Repubblica 2020, n.p.). The number of cases quickly rose, making Italy the number one hotspot for the disease. By the end of February, Italy had over a thousand confirmed cases and reported 29 deaths related to the virus (cf. Ananasso et al. 2020, n.p.). After the significant rise of cases in Italy the cases began to also increase in Spain, with a report of 50 cases by February 29 (cf. DSN 2020, n.p.). Both in Italy and Spain the number of cases rapidly rose from March onwards.
In order to contain the virus, the Italian government divided the affected regions into three zones. A red zone, that put the whole area in quarantine, a yellow zone where mainly y social events were suspended and the rest of Italy where public places were restricted (cf. Ministero della Salute 2020, n.p.). The measures in the red zones included, social distancing rules, such as keeping 1 metre to other persons, greeting other people only from the far and avoiding crowded places. People that were 75 or older had to stay home and people over 65 were obliged to stay home, if they had any sickness or health issue. All social events were cancelled, and schools were closed. Most university activities were also cancelled. However, students were allowed to finish ongoing exams and to perform artistical activities, such as music, wherever possible (cf. GU Serie Generale n.55 del 04-03-2020). By March 09, Italy declared a nationwide lockdown, since the number of deaths kept rising significantly. All people were ordered to stay at home and were only allowed to exit with a valid work, health or family reason. All sporting and entertainment events were completely cancelled, and all schools and universities were closed until April 3 (cf. BBC 2020, n.p.). Still, the number of total cases rose to over 35 000 by mid-March with Bergamo becoming the number one hotspot of the virus. With rising concern over the situation in Italy and having an increased number of deaths, the Spanish government declared a state of emergency on March 14. Spain was placed under lockdown, allowing citizens to only exit their homes to visit grocery stores, acquire pharmaceutical needs or visit financial and insurance institutes. An exception was also made for people who travelled to their workplace or had any health issues. The capacity of public transport was though reduced by 50 %. Schools and universities were also closed, and online classes

had to be provided. All nightlife institutes, small businesses, hotels and restaurants were also closed. All cultural events, as well as museums, libraries and monuments were also shut down. Religious institutions nonetheless remained open, while having to maintain containment measures, such as social distancing and hygiene rules.

By the end of March, the number of confirmed cases had risen to over a 100 000 in Italy (cf. Ruotolo 2020, n.p.) and to approximately 95,000 in Spain (cf. DSN 2020b, n.p.). The Spanish government extended the national lockdown until April 11 on March 26 (cf. Mayor Ortega 2020, n.p.). On April 09, the Spanish government extended it again until April 26 (cf. Castro 2020, n.p.). The lockdown was extended for a third time until May 10 on April 22 (cf. La Moncloa 2020, n.p.). Italy followed the same example extending the national lockdown to the period until April 13 on April 1 (cf. Cottone 2020, n.p.) and extending it again until May 03 on April 10 (cf. Biarella 2020, n.p.).

It is evident that both countries resorted to drastic measures in order to achieve the goal of flattening the curve during the first wave. These measures were adjusted and thus kept changing on a weekly basis. Both countries showed a similar strategy focusing on the isolation of individuals and affected regions. Nevertheless, the governments were forced to apply the national lockdown in order to constrain the rapid spread of the virus effectively. Table 1 shows a summary of the biggest measures applied during the first wave. The measures that will be part of the sentiment and emotion analysis in chapter 6 and 7 are highlighted in bold.

<table>
<tr><th></th><th>Italy</th><th>Spain</th></tr>
<tr><td>Jan</td><td>Jan 31
➢ Suspension of flights from and to China
➢ Isolation of infected individuals</td><td>Jan 31
➢ Isolation if infected individuals</td></tr>
<tr><td>Feb</td><td>Feb 21
➢ 10 municipalities are put under quarantine in Northern Italy
➢ Suspension of sporting and cultural events
➢ Areas are controlled by military and police</td><td>Feb 24
➢ 700 hotel guests are put under quarantine
➢ Travelling to China, Northern Italy, Japan, Iran and Singapore is “not recommended” by Health Ministry (cf. Gutiérrez 2020, n.p.)</td></tr>
<tr><td>Mar</td><td>Mar 1
➢ Affected regions are divided red and yellow zones
Mar 4
➢ Shutdown of schools and universities for two weeks (cf. Guerzoni et al. 2020, n.p.)</td><td>Mar 10
➢ Schools are closed for two weeks in Madrid, Vitoria, Labastida and La Rioja (cf. La Rioja 2020, n.p.)
Mar 12</td></tr>
</table>

	Mar 7 ➢ Lockdown in Northern Italy **Mar 9** ➢ Nationwide lockdown **Mar 22** ➢ all non-essential commercial activity is shut down (cf. GU Serie Generale n.76 2020, n.p.)	➢ Lockdown for four municipalities issued by Catalan government (cf. Baquero et al. 2020, n.p.). **Mar 14** ➢ State of emergency and nationwide lockdown is declared **Mar 26** ➢ first lockdown prolongation
Apr	**Apr 1** ➢ first lockdown prolongation **Apr 10** ➢ second lockdown prolongation	**April 9** ➢ second lockdown prolongation **April 22** ➢ third lockdown prolongation

Table 1 - Measures applied during the first wave

These measures showed great results combined with the participation of the public and the rise of temperatures.

3.2 Easement of restrictions from May 2020 to August 2020

With the help of strict containment measures, such as nationwide lockdowns, both countries began with the easement of restrictions after the number of detected cases began to decrease. Italian Prime Minister Giuseppe Conte announced the first easement plan on May 04. Italian citizens were allowed to travel freely within their regions and to visit relatives, provided they wear a mask. The movement between regions was still prohibited, except for work, health or emergency reasons. Face masks were also mandatory on public transport. Exercise, or a walk outside was allowed, providing a physical distance was maintained (cf. Ciriaco et. al 2020, n.p.). Further easements were announced on May 16, which allowed citizens to meet with friends, if they were not infected with the virus or had any related symptoms. The use of a mask was recommended for indoor activities and outdoors when a safe distance could not be kept. From 18 May onwards, commercial activities were restored, sports training was allowed to carry on and museums could reopen (cf. Guerzoni 2020, n.p.). Schools still remained closed until September. In Italy, the nationwide lockdown was ended on June 03, restoring free movement for all citizens (cf. Barone et al. 2020, n.p.).

The Spanish government could also ease the restrictions by the beginning of May 2020. The government announced a de-escalation plan on April 28, which consisted of the coordinated restoration of all activities and free movement in four stages with a duration of two weeks each.

In first stage, which started on May 04 and was called “phase 0”, children were allowed to exit their homes for one hour per day and adults were allowed to exercise outdoors. Restaurant deliveries were allowed, and many institutions could be visited for individual admission, upon appointment. Professional sports athletes could also resume their training. Phase 1 was applied from May 10 onwards and consisted of the restoration of social contact, the opening of small businesses and open-air markets. Hotels and restaurants could reopen under certain restrictions and hygiene rules and cultural and religious events could take place again. Two weeks after that, in phase 2, all businesses could reopen regardless of their size. Restaurants could provide tables but had to guarantee a certain distance between them. Educational institutes were reopened, as well as theatres and cinemas. In phase 3 all activities were reopened, while maintaining security measures and social distancing. Restaurants could use 50% of their capacities and discotheques and bars were also reopened (cf. Sánchez Hidalgo et. al. 2020, n.p.). The Spanish nationwide lockdown ended on June 21 (cf. Marcos 2020, cf.)

3.3 National containment measures during the second wave

After the application of the restriction easements, both countries, as did most European countries, were faced with rising case numbers from the end of July onwards. This caused both governments to apply new containment rules. Italy closed all discotheques and nightclubs on August 17 while the wearing of face masks became mandatory in public places until September 17 (cf. Voltattorni 2020, n.p.). The law was extended to the mandatory use of masks both indoors and outdoors on October 8, 2020 (Mari et al. 2020, n.p.). This excluded persons who were performing sports activities outdoors, as well as children and persons with disabilities. The decree also recommended the use of face masks on private meetings with persons, who were not cohabitants. Social distancing was also made mandatory, encouraging people to keep a distance from each other of at least one meter. Restaurants, bars and pubs had to close by 12.00 pm, if they offered table services and by 9.00 pm, if they didn’t. Home deliveries were not affected. Private parties were not recommended and only allowed a maximum of 6 guests. Religious institutions could carry on with their activities, if they maintained social distancing rules and kept a limit of 30 persons. Sporting institutions were kept open under certain containment rules and spectators were limited. Nevertheless, contact sports were forbidden entirely. Schools remained open and could carry on the school year with some restrictions, such as the prohibition of exchanges and school trips (cf. GU Serie Generale n.248 2020, n.p.). Nevertheless, having registered a significant increase of COVID-19 cases in mid-October, the containment measures were extended on October 25, 2020. The closure time for restaurants and bars was adapted to 6.00 pm, excluding Sundays. Table services were allowed for a maximum of four persons. This limit did not apply to cohabitants. All cinemas, theatres and discotheques were closed again. Private parties were forbidden, and all nightlife was suspended. Urban streets and Plazas could be closed by regional governments after 9.00 pm, if necessary. All ski resorts were closed, except for athletes and professionals. The free movement was not restricted, but it was strongly recommended to only exit your home for work, study, health and emergency reasons. It was also strongly recommended to limit personal visitations and travel. About 75% of Italy's education institutes had to apply online-education methods. All fitness centres and pool resorts were closed, and all sporting events had to take place without spectators. Most commercial and industrial businesses could remain open. Protests and marches were not forbidden, as long as all containment measures were applied and they took place in the form of static demonstrations or sit-ins. With cases still increasing and the second wave still ongoing in the beginning of November, Italy released a new decree on November 3, 2020. The government reintroduced the division of regions into red (level 4) and orange zones (level 3) depending on their individual situation. A lockdown of at least 15 days was imposed in all red zones. All non-essential commercial businesses were closed. All night-life activities were halted, except for food delivery services. Elementary schools and kindergartens remained open. For all orange zones the displacement to other regions was forbidden, except for health, emergency, study or work reasons. All restaurants, pubs and bars were closed, except for canteens and catering services.

Different nationwide measures were also applied. A curfew from 10.00 pm onwards was introduced, except for work, health or emergency reasons. All Museums were closed and all higher education institutes had to apply online-methods. An exception was made for students and researchers, who had to work in laboratories. All kindergartens, middle schools and high schools remained open, but a mandatory use of masks was imposed. All non-essential commercial businesses were closed. Public transport was limited to 50 %. All bars and restaurants closed at 6.00 pm but could remain open on Sundays. For all citizens it was recommended to limit visitations and to only travel for work, study and health reasons (cf. G.U. Serie Generale , n. 265 del 25 ottobre 2020). With a new decree released on December 3, strong containment measures were applied for the festive period. Travel between regions was forbidden during the festive period from December 21 to January 6, except for work, health or emergency reasons. On December 25 and 26 and on January 1, it was also forbidden to travel within the region. On New Year's Eve a curfew was applied from 10.00 pm to 7:00 am (cf. GU Serie Generale n.301 del 03-12-2020).

The summer of 2020 had particularly affected the situation of the pandemic in Spain.

In August 2020, Spain had become Europe's epicentre for the virus with an incidence rate of 189,6 per two weeks, which was 10 times higher than Italy's rate at the time by October 30 (Hernández 2020, n.p.). With the decree from August 14 the government decided to close discotheques, dance halls, pubs and bars. Restaurant visitors had to keep a distance of at least 1,5 m. Table services were allowed with a limit of 10 persons per table. All restaurants and food service establishments had to close by 1.00 pm. The consumption of alcohol in the public was prohibited. All affected regions were also heavily tested with PCR tests in order to identify the infected persons (cf. Ministerio de Sanidad 2020, n.p.). With most of the detected cases coming from the capital, new containment rules were applied to different municipalities of Madrid on September 18. Traveling within other regions was prohibited, except for work, health and study reasons and to go to banking, legal and administrative institutions. Private reunions were limited to six persons, except for cohabitants. Religious institutions remained open with a limit of 10 persons indoors and 15 persons outdoors. All commercial businesses had to close at 10.00 pm, except for food delivery services. Parks, recreation centres and gardens were closed entirely. Sports institutes remained open with a limit of 50 % of visitors and with a maximum of a six-person group. Private reunions were limited to six persons within Madrid (cf. Valdés 2020, n.p.). After a small downfall of registered cases by the end of September the numbers began to rise again in the beginning of October. The Spanish government reacted by imposing a partial lockdown on municipalities with over 100 000 residents, with 500 cases per 100.000 inhabitants, 10% positive PCR tests and 35 % of intensive care units occupied on October 1. Most affected municipalities were still in Madrid. Still, the numbers kept rising and Spain became the first European nation to reach 1 million COVID-19 cases by mid-October (cf. BBC 2020b, n.p.)

On October 25 the government reintroduced the state of emergency. The decree established legal coverage for all restrictions applied in Spain. A curfew was applied from 11.00 pm to 06.00 am, except for work, study, health and emergency reasons. An exception was made for the Canary Islands, because of their low infection rate. Travel within regions was prohibited and it was strongly recommended that citizens limit their movement and stay at home whenever possible. Private gatherings were also limited to six persons of different households. All measures were applied for a period of 15 days and could be modified by regional governments (cf. Munera 2020, n.p.) The state of emergency was extended for six months on November 9. Gatherings for Christmas and New Year's Eve could take place within a limit of up to six people, if a family member was not a cohabitant and with a limit of 10 persons, if all members lived in the same household. Religious festive events could take place without singing. Christmas markets were allowed to open, if security measures were maintained. A nightly curfew was also established, limiting mobility from 1 am to 6 am. On New Year's Eve a curfew was established to 1 am, allowing citizens an hour to celebrate, including the return to their home. It was also recommended to avoid meetings and celebrations of multiple persons and to avoid unnecessary travel (cf. Lamet 2020, n.p.).

The summary of the measures applied by both countries shown in Table 2 demonstrates a slight difference of both countries during the handling of the second wave. Italy reacted more quickly and actively to combat the rising numbers and applied national rules as early as in August. The Spanish government started with limiting the restrictions to affected regions but were obliged to reinstate the state of emergency at a later stage as cases kept rising. Italy had a more passive approach during September, but applied stricter rules during the festive period.

	Italy	Spain
Aug	**Aug 17** ➢ Discotheques and nightclubs are closed ➢ Mandatory face masks indoors until September 17	**Aug 14** **Containment measures only within affected regions** ➢ Discotheques, dance halls, pubs and bar are closed ➢ Restaurants open with social distancing ➢ Table services allowed with limit of 10 persons/table ➢ Restaurants/food service to close by 1.00 pm. ➢ The consummation of alcohol in public prohibited

		➢ All affected regions are heavily tested with PCR tests
Sep		**Sep 18** **Restrictions for municipalities in Madrid** ➢ Private reunions were limited to six persons (except cohabitants) ➢ 10 (outside) / 15 (inside) person limit for religious institutions ➢ Commercial business had to close on 10.00 pm (except food delivery services) ➢ Parks, recreations and gardens were closed ➢ Limit of 50% of visitors for sports services (max. 6 persons) ➢ Limit of 6 persons for private reunions
Oct	**Oct 8** ➢ Mandatory face masks indoors and outdoors ➢ Recommendation of use during private meetings ➢ Mandatory Social distancing ➢ Restaurants, bars, pubs close on 12.00 pm with table services / 09.00 pm without tables ➢ Private parties prohibited ➢ Limitation of religious gatherings to 30 people ➢ Contact sports prohibited **Oct 25** ➢ Adaption of closure times for restaurants, bars and pubs to 06.00 pm (Sundays excluded) ➢ Table services limit to 4 persons ➢ Cinemas are closed ➢ Ski resorts, fitness centres, sports resorts are closed for amateur athletes ➢ Sports events take place without spectators ➢ Recommendation of restriction of movement and visitations ➢ Protests marches allowed only as static demonstrations or sit-ins. ➢ Online education for 75% of education institute members	**Oct 1** **Partial lockdown** for: municipalities with over 100 000 residents, with 500 cases per 100.000 inhabitants, 10% positive PCR tests and a 35 % of intensive care units occupied. **Oct 25** **Declaration of state of emergency** ➢ Curfew from 11.00 pm to 06.00 am (except for work, study, health and emergency reasons) ➢ Prohibition of travel within regions ➢ Recommendation of limitation of movement private gatherings limited to six persons of different households.
Nov	**Nov 3** **Division of regions into red (level 4) and orange zones (level 3)** *Red zones:* ➢ Lockdown for all red zones (at least 15 days) ➢ Closure of non-essential commercial businesses and industries	**Nov 9** Extension of state of emergency for six months Christmas and New Year's Eve restrictions decided

	➢ Closure of nightlife services (except food delivery) *Orange zones:* ➢ Prohibition of travel to other regions (except for health, emergency, study or work reasons) ➢ Restaurants, pubs and bars were closed (except for canteens and catering services) *Nationwide measures:* ➢ Curfew from 10.00 pm onwards (except for work, health or emergency reasons) ➢ Museums closed, online education for higher education (except for laboratory activities) ➢ Mandatory masks for kindergartens, middle schools and high schools ➢ non-essential commercial businesses were closed. ➢ Public transport was limited to 50 %.	
Dec	**Dec 3** ➢ Prohibition of travel between regions during the festive period (Dec 21 – Jan 6) except for work, health or emergency reasons. ➢ Prohibition of travel within region on Dec 25 and Jan 1 ➢ Night curfew from 10.00 pm to 7.00 pm on New Year's Eve.	**Restrictions on Christmas and New Year's Eve** ➢ Limit of six persons, if a family member was not a cohabitant ➢ limit of 10 persons, if all members lived in the same household ➢ nightly curfew from 1 am to 6 am. ➢ New Year's Eve curfew from 1 am onwards, Recommendation of avoiding groups of people and big celebrations

Table 2 - Measures applied during the second wave

The pandemic required a constant adaptation of the containment rules that affected the lifestyle of the population of both countries. The reaction to rising numbers and the restriction rules were a heavily discussed topic both in the political sphere, as well as the social sphere. The population of both countries used social media and new sites to express themselves and share their own opinions and views on the matter. Consequently, these online sources are great options to base the sentiment and emotion analysis. Before applying the analysis method, the next chapter will aim to give an overview of sentiment and emotion analysis as a classification task and emotion theories that are commonly used as a classification basis.

4 Defining sentiment and emotion

Sentiment and emotion analysis are widely used terms that can convey different meanings according to the referring research or professional field. One of the first works on the topic was released by Janyce M. Wieber as early as 1990 in the paper *Identifying Subjective Characters in Narrative.* Wieber built one of the first algorithms that tackled the problem of identifying the viewpoint of certain fictitious characters in novels by creating an according algorithm (cf. Wieber 1990, 1f.). Since then, the task of analysing sentiment and emotion expressed in text has steadily gained popularity and has become an important application of natural language processing and other research fields. Other than rule-based approaches, both machine learning and deep learning have contributed to a new rise of research interest in the field, while social media and web data have provided broader opportunities for data analysis.

4.1 The variety of sentiment analysis research

The term sentiment analysis is commonly used to describe the process of extracting and analysing the sentiment or opinion about a certain subject or entity. The analysed data can range from pictures, videos, audio and text sources. Nonetheless, most research studies focus on the extraction of sentiment in written text. This is the case for the research field of natural language processing, from which most studies are published. Nevertheless, the task of analysing sentiment has become more prominent in other fields, such as information retrieval and data mining, as well as financial studies and linguistics. Recent studies also include the aspect of multimodality to obtain a more detailed analysis (cf. Liu 2020, 1). The progress in research has been facilitated by the rise of social media, blogs and other web communities, which help provide a vast quantity of data that can be used for examination. These platforms have accordingly become the focus point of researchers, since its value lies in the expression of opinions and feelings towards certain topics (cf. Liu 2020, 3f.). Apart from the textual data, many social media platforms provide other social functions, such as likes and dislikes, ratings or emotional reactions using emoticons, which can be included in the analysis process.

The opportunity of social media is exploited to perform a variety of applications. The term sentiment analysis is thus used for difference tasks that are often related to each other. It can be generally described as the process of automatically analysing the sentiment, appraisal or attitude expressed in a document. However, the term can be also used for related applications, such as the determination of opinions (opinion analysis or opinion extraction) or multimodal sentiment analysis (cf. Liu 2020, 1). Nonetheless, the terms opinion mining and sentiment analysis are used interchangeably in academic research studies. This is due to a lack of general interest in defining the terms “sentiment” and “opinion” beforehand due to the subtle difference in the definition of the terms. Cambridge dictionary defines both terms as the following:

> **Sentiment:** a thought, opinion, or idea based on a feeling about a situation, or a way of thinking about something (Cambridge Dictionary).
>
> **Opinion:** a thought or belief about something or someone Cambridge (Cambridge Dictionary).

The definitions show that the terms are related and can be dependent on each other. While an expressed sentiment can imply an opinion about a certain topic, a sentiment can be caused by an opinion about a topic. Nevertheless, the above definition also shows, that the term “sentiment” implies a feeling towards an entity, whereas the term “opinion” generally means a view about an entity (cf. Liu 2020, 2f.).

This study will exclusively use the term sentiment analysis to describe the analysis of the feeling expressed in the analysed text. The use of the term will imply the analysis of the polarity in the examined documents, which will be narrowed down to the classification as positive or negative. The aspect of multimodality, which includes the images, videos and audio published by the news sites, will not be taken into consideration, since this would exceed the time frame of this thesis.

The performance of sentiment analysis generally depends on the complexity of the approach and the extent of the analysed aspects. In general, three different levels can be used: the document level, sentence level and the aspect level. The document level analysis assumes that a document expresses a feeling about a single entity (e.g., film review) and determines the negative or positive sentiment contained in a complete document. The sentence level analysis distinguishes subjective opinions and sentiments about a topic and returns a positive, negative or neutral output. The aspect level analysis is a more fine-grained analysis level and focuses on specific target entities (cf. Liu 2020, 10f.).

The applied method can vary as well, depending on the industrial or research preference. Most analysts use machine learning methods, dictionary-based approaches, or a combination of both (hybrid method). For the dictionary approach a list of sentiment words is used to determine the sentiment expressed in a document. Each sentiment expression is detected and marked with a sentiment score usually between -1 and +1. Many approaches also take negations and other syntactic rules into consideration. A sentiment function is then applied to calculate the aggregate score (cf. Liu 2020, 122f.).
The machine learning method defines the task as a classification problem. An algorithm is trained with labelled data and determines the sentiment of new data as positive or negative by using a classification method, such as SVM or Naïve Bayes. Deep learning methods can also be used to perform sentiment classification. After text processing and normalisation, the raw text input is vectorised to create numeric tensors (cf. Kolchyna et al. 2020, 20). The approach used in this study will be based on a Convolutional Neural Network (CNN), which is typically used in Deep Learning tor tasks, such as image and video classification. A CNN typically uses a tensor as its input, which goes to a series of processing layers and returns a new tensor (cf. Kolchyna et al. 2020, 20).

4.2 Emotion theories and their role in emotion detection

The role of emotions in our everyday life is of undisputable importance since they prepare and help us deal with occurrences and react to stimuli within our immediate environment. Consequently, emotions have been thoroughly analysed in different fields, such as psychology, philosophy, and neuroscience. The study of emotion has nonetheless recently gained prominence in linguistics and computational linguistics. In computational linguistics and data science the determination of emotion is a branch of sentiment analysis that specifically deals with the classification and extraction of emotion in a target document.
Emotion detection is usually carried out with the help of a specific model, in which emotions are categorised. The categorisation of emotion is a complex research field and has been the focus of many studies that use different methods and theories. These theories can be classified into four major traditions. The Darwinian and evolutionary tradition determined that emotional expression can be tracked back to the evolution process. In his work *The Expression of the Emotions in Man and Animals (1872)* Charles Darwin studied

and collected animal expressions as well as own animal observations (cf. Plutchik 1980, 1f.). Darwin sees emotion in its functional role, serving as signals and preparation for certain actions (cf. Plutchik 1980, 3f.). His work also provided evidence for the existence of innate emotions that appear in animals, as well as very young children. In addition, he determined that certain emotions are innate, regardless of blindness and distinct races and social groups (cf. Plutchik 1980, 5).

The second major emotion theory was founded by William James. James presented a theory that contradicted the popular belief that physical changes appeared as consequence of feeling certain emotions. Instead, he proposed an inverted definition, claiming that physical changes preceded emotions. He thus claimed that emotions were to be defined as the feeling of certain bodily changes caused by all parts of the organism and applied this theory to “basic emotions”, such as grief, fear, rage, and love, while excluding more complex emotions (cf. Plutchik 1980, 6ff.). Since the Danish physiologist Carl Lange proposed a similar idea, this theory known as the James-Lange theory (cf. Plutchik 1980, 9). The James-Lange theory was discredited by the physiologist Walter B. Cannon a few years later in the book *Bodily Changes in Pain, Hunger, Fear and Rage* published in 1915. The author separated the sympathetic division of the autonomic nervous system and thus abolished the vasomotor centre in several cats and determined that the animals showed emotional reactions, such as fear, rage, and pleasure. This showed that there was no relationship between emotional expression and the artificially caused lack visceral changes (cf. Plutchik 1980, 11). Furthermore, he pointed out that stressful stimuli cause the same physical reactions. In addition, viscera could be damaged without producing any discomfort. Due to the lack of quickness of the small portion of information by the viscera it is unlikely that a human would recognise his emotional state through his bodily changes. Cannon also pointed out that the artificial inducement of visceral changes, that were typical for strong emotions did not cause any emotional feelings (cf. Plutchik 1980, 11).

Cannon instead proposed a different theory, which determined that emotion depends on neural discharges based on the optic thalamus. These discharges produce an emotional experience, as well as a bodily charge simultaneously (cf. Plutchik 1980, 13).

Lastly, the Freudian dynamic tradition is mostly based on subconscious emotions and their psychoanalytic meaning. It claims that both the feeling of emotion and the physical reaction both result from an evaluation of the subconscious and thus eliminates the sequence problem (cf. Plutchik 1980, 19 f.). Most studies on emotion detection and analysis apply emotion theories that extend the Darwinian evolutionary tradition. The most prominent models used in emotion detection are Ekman (1992), Izard and Plutchik, who use a categorical method to determine emotions. These authors postulate a set of basic emotions, which can be extended to a more complex range of emotions. With happiness, surprise, sadness, fear, disgust, and anger Ekman considers six basic emotions (cf. Ekman 1992, 170 f.). Plutchik's model extends Ekman's basic emotions, adding anticipation and trust (cf. Plutchik 1980, 160) and Izard's model also includes guilt and shame (cf. Izard 1977, 46). Nonetheless recent studies have implemented cognitive models, such as Scherer's model (1999), which views emotion as a reaction to people's appraisal of an object or event (cf. Roseman et al. 2001, 3).

These theories are though only a small proportion of the most cited works in emotion detection papers. Many other approaches have been used to describe emotions in other fields, such as psychology, philosophy and linguistics. The broad range of emotion theories is consequently accompanied by a dissension of the definition of the word.

The definition used in this study will be based on Plutchik's psychoevolutionary theory of emotion, since it covers a broader range of emotions. Ekman's theory focuses mainly on negative emotions, while Plutchik also integrates positive ones. Furthermore, Plutchik's theory postulated that emotions are a survival response to the immediate environment, which can be applied to the analysed context.

4.3 Robert Plutchik 's psychoevolutionary theory of emotion

Robert Plutchik's emotion theory derives from Darwin's evolutionary perspective and aims to categorise emotions into basic primary innate emotions and secondary complex emotions that result from the combination of those. The author defines emotion as follows:

> "Emotion is an inferred complex sequence of reactions to a stimulus, and includes cognitive evaluations, subjective changes, autonomic and neural arousal, impulses to action, and behavior designed to have an effect upon the stimulus that initiated the complex sequence. […]." (Plutchik 1982, 551).

Robert Plutchik's definition integrates the complexity behind the formation of emotions and built the foundation for a dimensional categorisation of emotion. His theory provides a structural model and a possible procedure for the development of complex emotions with the use of eight basic emotions: anger, joy, fear, trust, surprise, anticipation, trust, and disgust (cf. Imbir 2017, 1). His Psychoevolutionary Theory of Emotion is based on 10 postulates:

> 1. The concept of emotion is applicable to all evolutionary levels and applies to animals as well as to humans.
> 2. Emotions have an evolutionary history and have evolved various forms of expression in different species.
> 3. Emotions serve an adaptive role in helping organisms deal with key survival issues posed by the environment.
> 4. Despite different forms of expression of emotions in different species, there are certain common elements, or prototype patterns, that can be identified.
> 5. There is a small number of basic, primary, or prototype emotions.
> 6. All other emotions are mixed or derivative states, that is, they occur as combinations, mixtures, or compounds of the primary emotions.
> 7. Primary emotions are hypothetical constructs or idealized states whose properties and characteristics can only be inferred from various kinds of evidence.
> 8. Primary emotions can be conceptualized in terms of pairs of polar opposites.
> 9. All emotions vary in their degree of similarity to one another.

Using those postulates the author derived six claims describing the concept of emotion. Firstly, he claimed that emotions are communication and survival mechanisms that increase the chance of survival by adapting individuals to the immediate environment using an appropriate reaction. His second claim is that emotions have a genetic basis, which provides a physiological mechanism to arbitrate behaviour. Thirdly, he postulates that emotions are hypothetical constructs whose characteristics can only be determined with evidence. Furthermore, he proposes that emotions are a complex chain of events that produces behavioural homeostasis. Through cognitive processing, the feeling of emotion and the physical changes occur at the same time. His fifth claim is that emotions can be represented in a three-dimensional model, which represents the intensity of each emotion and the relationship between those emotions. Lastly, he postulates that emotions derive from multiple concepts. The language of emotion structure can be found in personality traits or ego-defense mechanisms. These concepts can be portrayed in a circle model that links primary and secondary emotions. Using these claims Plutchik developed an emotion circle, which is known commonly referred to as Plutchik's wheel of emotions (cf. Imbir 2017, 2f.). Plutchik placed the primary emotions are placed on a wheel (see figure 2) with heir vertical secondary states and third states represented the according emotion intensity (cf. Plutchik 1980, 157).

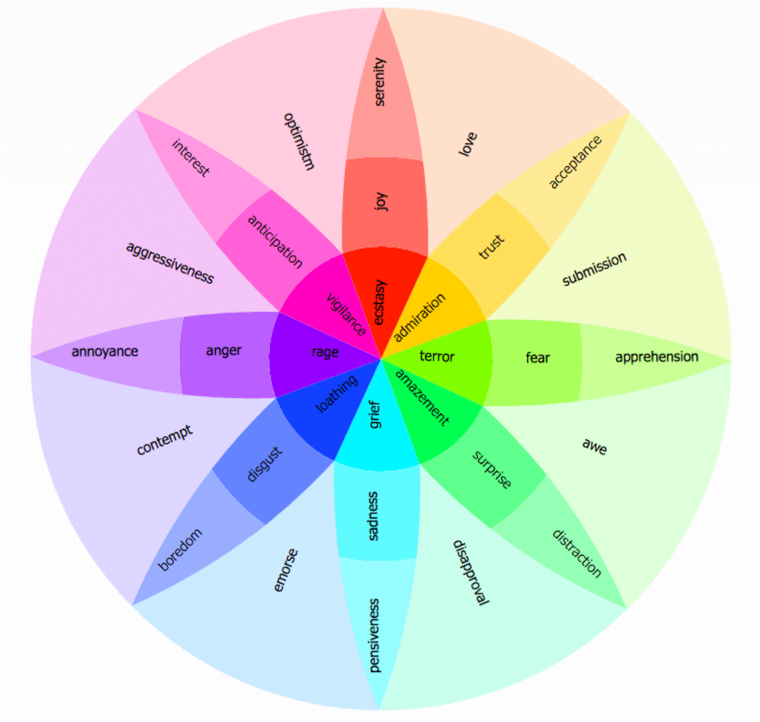

Figure 2 - Robert Plutchik's wheel of emotion[8]

Furthermore, each emotion is paired with a polar opposite, which is based on the physiological reaction to a certain event (cf. Plutchik 1980, 164):

- Joy is the opposite of sadness.
- Fear is the opposite of anger.
- Anticipation is the opposite of surprise.
- Disgust is the opposite of trust (cf. Plutchik 1980, 164).

Moreover, the wheel of emotions represents the links of primary and secondary emotions, which results in new complex emotions. Adjacent pairs of primary emotions are combined into complex intermediate emotions. The mix of joy and trust results in a complex emotion, such as love and/or friendliness, for example (cf. Plutchik 1980, 160-163.).

[8] Source: Nielek, Radoslaw & Ciastek, Miroslaw & Kopeć, Wiesław. (2017). Emotions make cities live. Towards mapping emotions of older adults on urban space.

4.4 Summary

The use of the terms sentiment and emotion analysis are often used for different applications in different research areas. Both tasks are often performed as classification problems and use rule-based approaches, as well as Machine Learning and Deep Learning. While the classification of sentiment analysis is mainly based on a three-level basis of positive, negative and neutral, emotion detection bases its classification on motion models used in psychology and neuroscience. Primary “basic” emotions determined by Paul Ekman, Plutchik or Izard often build the basis for the categorisation used in the emotion detection algorithms. New research papers also use cognitive models derived from neuroscience.

Most studies on sentiment analysis and emotion detection base their analysis on textual data extracted from social media, while also integrating interactive social functions in the analysis. Therefore, most datasets and corpora are based on data obtained from Twitter and take little to no consideration of online news sites, which contain a great amount of data in the form of the comment section. In addition, most research papers and data sets are based on the English language. This study aims to fill this research gap by building two news corpora, that can not only be used for sentiment analysis, but also for a variety of other linguistic and data science tasks. The corpora contain textual and meta data found in articles about COVID-19 in the newspapers El Mundo and La Repubblica.

5 Creating the press corpora

The newly created corpus is based on a collection of press data from the Spanish newspaper *El Mundo* and the Italian newspaper *La Repubblica*. Using the corresponding web pages of each newspaper, the aim was to collect as much information as possible about the articles about COVID-19.

For this purpose, a two-step process was applied consisting in the collection of the links of the targeted news articles and its comments and the storing of the data in a structured format that can be used for further analysis. This task was performed automatically by building a python-based web crawler. The number of articles collected varied depending on the output of each newspaper. The collection of the available data required a thorough analysis of the web pages in their structure and by inspecting their source code. Most web pages contain a search engine that can be used to search for articles about certain topics. In addition, many news websites use their own article archive, which can be further used for a more advanced search. The collection of the data was performed by using web scraping. Web scraping consists of accessing multiple websites and collecting the information provided by them. While this process can be performed manually, it is generally performed by building an automated programme, also known as web crawler, that interacts with a server, requests the data of multiple webpages, and parses it to extract the needed information (cf. Mitchell, 2015, p. 7).

Web crawlers are usually used in web search engines and other systems that assemble a corpus of certain pages and assign them an index that can be later found by user queries. Many web pages are also archived in a structured way for later access. Web crawlers are also increasingly used for data mining and web monitoring services (cf. Olston et al. 2010, p. 176). Using a web crawler for each news website, the data could be collected automatically and saved in a local file. The first step was to collect all articles related to COVID-19. After that, all articles were accessed separately in order to obtain the necessary information. The extracted data was then saved as an XML file as well as TXT files for each extracted article.

5.1 Searching for COVID-19 articles

The first step of the crawling process consisted of the search of targeted articles. For this purpose, the crawler needed to access the website and enter the selected keyword in the search field or the web archive. For this study, the keyword “coronavirus” was chosen, as other keywords, such as “COVID” or “COVID-19” returned a significantly smaller result. The search was performed using the Python library Selenium[9]. The Selenium library, which is mainly used for automatic website testing, was used in this case as a convenient tool for web scraping tasks, especially when the websites used JavaScript or Jquery for their functionality. Using this library, the websites could be accessed as they are portrayed in a standard web browser. Almost any popular browser, such as Google, Firefox or Microsoft Edge contains integrations that allow the Selenium library to automatically load websites, perform certain actions, take screenshots, and save the required data automatically (cf. Mitchell, 2015, p. 214). The crawler built for this study was used to simulate the opening of the browser, localise the search field, click on it and enter the selected keyword. The action is then performed by simulating the operation of the “Enter” button or with a simulated click on the required button. Both webpages used a search function that could be adjusted according to the needs of this study (see figure 3).

[9] Selenium is a library for automated browser tasks, mainly used for testing purposes (https://www.selenium.dev/).

Creating the press corpora

Figure 3 - Search results for "coronavirus" - La Repubblica

The web pages returned a list of all articles that were published between January and December 2020 and contain the keyword “coronavirus” either in the title, tag, subheading or in the article text.

The crawler then inspected each result using the specified element tags used in the source code of the pages. The article links returned by the search query were analysed further. Using the Python library BeautifulSoup[10] the link titles could be accessed, and each URL was extracted. The extracted URLs were then stored in a CSV file for later access (see figure 4). For later categorisation the scraping process was divided by month.

[10] BeautifulSoup is a useful library that helps to extract information from HTML and XML data. It is one of the most popular libraries used to parse HTML, since it provides ways of navigating, searching and modifying the parsing trees (https://www.crummy.com/software/BeautifulSoup/bs4/doc/).

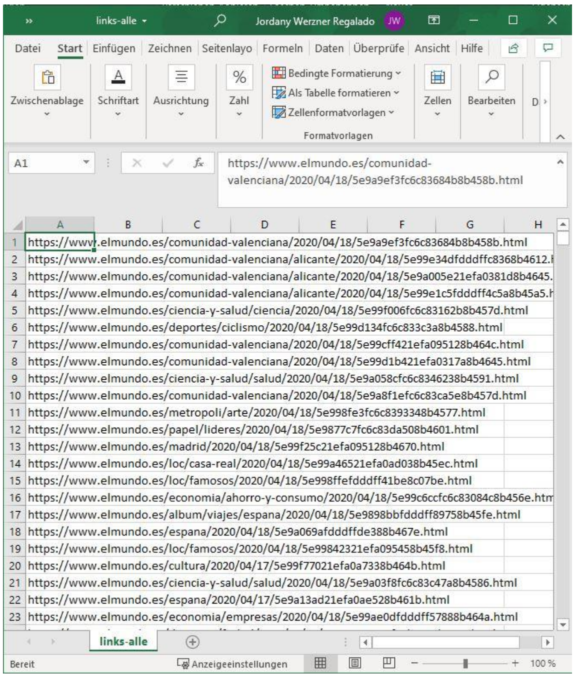

Figure 4 - Extracted article links - El Mundo

5.2 Information extraction – Scraping the news articles

For the extraction of the information each article link was accessed by the web crawler using the according CSV file. The Selenium library was used to open the articles in the web browser. The HTML data is then parsed using the Python library BeautifulSoup. The HTML content is transformed into a BeautifulSoup object, and its tags can be easily targeted and accessed. This allows the crawler to parse the data of each article and extract the needed information.

The crawler thus accessed the meta data of each article by searching for certain HTML elements and their classes (see figure 5).

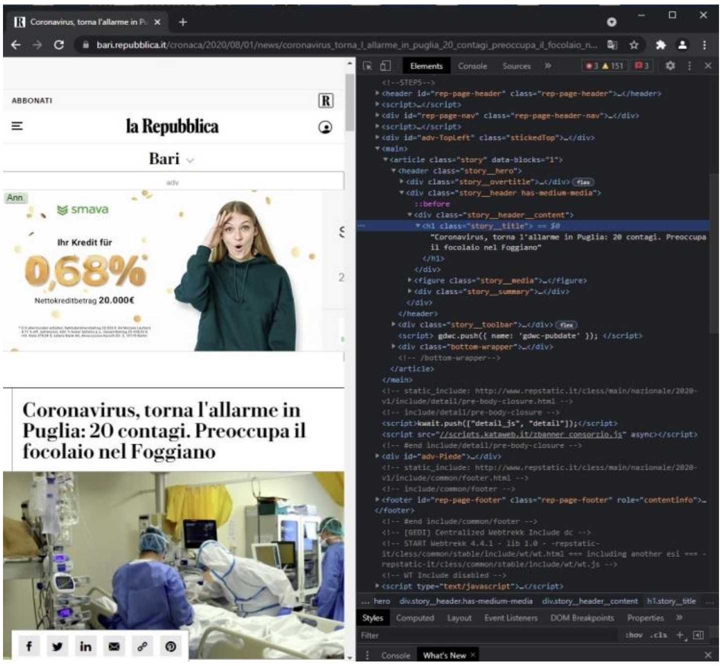

Figure 5 - Article source code - La Repubblica

Most articles were structured in a similar way so that the crawler was built to perform the same tasks. This was the case for headlines which usually contained the HTML tag *<h1>* and an appropriate class name. The subheadings were mostly published with the HTML tag *<h2>* or *<h3>*, while the article text was usually saved under the *<p>* tag. The article images were found using the *<img>* or *<figure>* tag. Other information such as the author or the data were stored using different *<div>* or *<p>* elements and required a more detailed analysis in order to be processed correctly. In multiple cases a different web page structure was found, and the crawler was updated accordingly by adding new tag and class names into the code structure.

For the extraction of the information of the comment section the crawler switched back to Selenium since most actions were based on JavaScript. Every information found in each comment, such as the author alias, comment date and comment text, also stored in HTML-tags, could also be accessed using the Selenium library. The information was saved in a variable for later processing. All actions were performed in a loop, until all articles were scraped and closed. Finally, the extracted data was then saved in an XML-format.

5.3 Saving the Data – Creating XML files

The prospective analysis of the extracted information was facilitated by creating an XML file for each article. The markup language XML is used to encode documents that contains information that is easily readable by humans and keeps a structure that can be processed by machines (cf. W3C, n.p.).

The XML used to store the information of the files can be seen in figure 6. All variables containing the extracted information were used to create the XML file.

```xml
<?xml version="1.0" ?>
        <?xml-model href="" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
        <TEI xmlns="http://www.tei-c.org/ns/1.0">
          <teiHeader>
            <fileDesc>
              <titleStmt/>
              <publicationStmt>
                <authority/>
              </publicationStmt>
              <sourceDesc>
                <listPerson>
                </listPerson>
              </sourceDesc>
            </fileDesc>
          </teiHeader>
          <text>
            <body>
              <div type="newspaper">
                <post type="article" auto="false" who="{who}" when="{when}" ref="{x}">
                  <head>
                    <headline>{headline}</headline>
                    <subheading>{subheadline}</subheading>
                  </head>
                  <image xlink:href="{image}"
                    xlink:title="{img_title}"></image>
                  {body}</post>

                <post type="comment" auto="false" who="{comment_who}" replyTo="{comment_replyTo}" ref="{ref}">
                    <p>{comment}</p>
                </post>
              </div>
            </body>
          </text>
        </TEI>
```

Figure 6 - XML file structure

The XML file format allows a separate analysis of the scraped information. Using the XML tags, such as headline, subheading, image, etc. or any XML attribute, the scraped data could be accessed separately. This facilitated the creation of TXT-files containing individual information, such as all headlines, all comments, etc.

5.4 The news corpora – A quantitative overview

Overall, a total of 32750 articles were collected using the crawler. As seen in Table 3 the numbers of scraped articles varied depending on each month.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
El Mundo	83	379	3470	3524	2543	1931	1406	1210	1578	1852	1404	1242	**20622**
La Repubblica	122	698	2405	2500	1617	992	715	599	499	775	706	500	**12128**

Table 3 - Number of extracted articles about coronavirus

Since El Mundo generally released more articles in total, more articles about the pandemic could be scraped. However, both newspapers show a very similar course of the articles released per month both having a spark of articles released during the first wave between March and May and a downfall between June and August during the easement of restrictions, before reaching a new spike during the second wave from September onwards. This dynamic similarity can be seen by plotting the table above, as shown in figure 7. A notable increase of publications is thus shown, depending on the gravity of the situation.

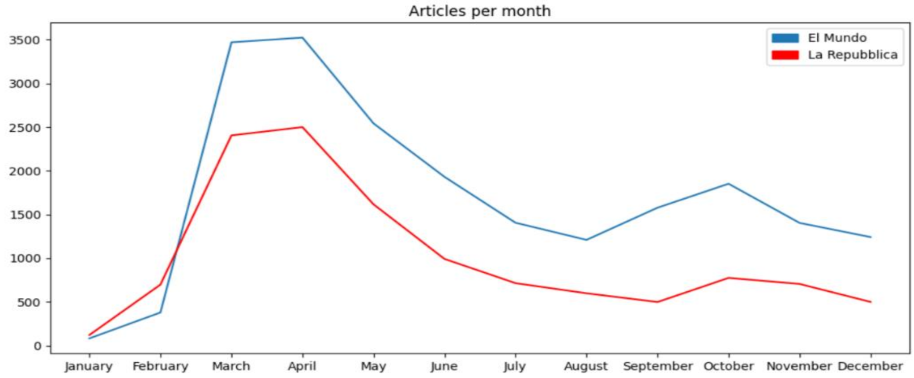

Figure 7 - Number of extracted articles about coronavirus (graph)

Over all 704 738 comments were extracted from El Mundo, while 8859 comments were extracted from La Repubblica. El Mundo thus contained a significantly higher number of comments. The main reason for this disparity is the different access of the comment section function. El Mundo only requires a small registration process, which can be done using Facebook or your Google account with one click, while La Repubblica reserves its comment section to paying customers. This disparity should be taken into consideration when analysing the results obtained by sentiment classification and emotion analysis.

6 General Sentiment Classification

Since the extracted meta information was not annotated beforehand, a Deep Learning approach was used to perform general sentiment classification.

The algorithm was built and trained with manually classified training data containing positive and negative comments obtained from Twitter data sets. The classification algorithm was built using Python's spaCy[11] library, which contains trained models and pipelines for the Spanish and Italian language. Using spaCy's pipeline the algorithm was trained to return a positive or negative output for the untrained data. Given the structure of the XML corpora, the meta data used for sentiment analysis could be individually selected using the according XML tags. Since it was expected that the article text generally contained a neutral language, the analysis was based solely on the comment section to portray the general sentiment of the readers of both newspapers by month. The corpora were filtered accordingly.

Furthermore, the sentiment was put into context by analysing the shifting sentiment during different chosen time frames. This approach was used to determine whether there was a relation between the applied government measures and the change of sentiment.

6.1 Text processing and normalisation

Before applying the chosen classification method, the text was pre-processed. The first step was to remove any sentences that appeared more than once in each corpus. This eliminated any spam comments or comments that were posted automatically by bots.

Any characters longer than 23 characters were also eliminated from the Spanish comments and those longer than 26 from the Italian comments. The character limit was chosen according to the longest Spanish word accepted by Real Academia Española and the longest Italian word found in Dizionario Garzanti[12]. This eliminated any typing errors or unusually long characters purposely used in the comment section.

---

[11] SpaCy is an open-source library that can be used to perform NLP tasks. It is published under the MIT license and focuses, unlike NLTK, on production usage. It can be used to build machine and deep learning models to perform classification tasks (cf. Honnibal et. al. 2017, n.p.).

[12] The longest word accepted by Real Academia Española (RAE) is "electroencefalografista" and the longest Italian word is "precipitevolissimevolmente" according to Dizionario Garzanti. However, for both languages longer words exist that are used in a scientific context.

Other text processing steps were performed using SpaCy. SpaCy contains pre-trained pipelines that, depending on their size, contain binary weights as a tokenizer, a POS-Tagger, a dependency parser, and a named entity recognizer to predict textual annotations. Furthermore, the pipelines contain lexical entries in the vocabularies, such as their independent attributes, as well as functions for text normalisation. Finally, spaCy creates word vectors that can be further used for Machine and Deep Learning tasks.

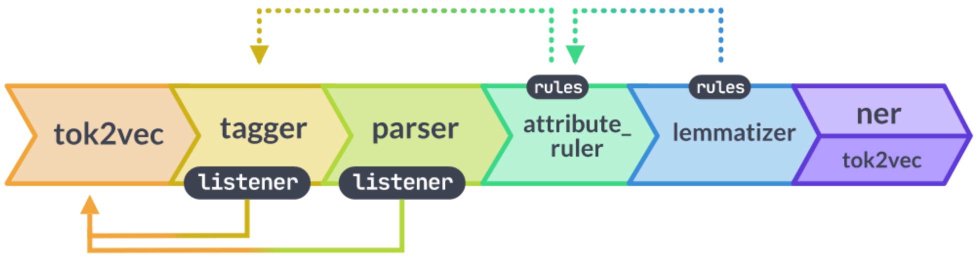

Figure 8 - SpaCy's text processing pipeline

Invoking the pre-labelled training comments into the spaCy pipeline thus results in a doc object that contains span and token objects of the source text (see figure 8). After tokenisation, the text is assigned part-of-speech tags with the tagger, dependency labels, named entity labels, the base forms of the tokens (lemmatisation) and document labels. The doc object thus contains a sequence of tokens and all its annotations (cf. Honnibal et. al. 2017, n.p.).

6.2 Using a convolutional neural network for sentiment classification

Convolutional neural networks usually take tensors as input, which go through different processing steps, which are called layers. A CNN runs layer by layer a forward pass. The input goes through the first layer, generating a new output. The output then goes through the next layer. When the processing through all layers is finished the final output is typically run through an additional layer, which is used for backward propagation. Finally, which measures the discrepancy between the prediction and the target (cf. Wu 2017, 5). Nevertheless, a single layer CNN architecture can be used to obtain great results, using pre-trained word vectors that were trained on a very large text corpus (cf. Kim 2014, 5). For this purpose, the algorithm was built using the Python library spaCy, which contains pre-trained language models, for the Italian and Spanish language (cf. Honnibal et. al. 2017, n.p.). The spaCy pipelines *it_core_news_lg* for the Italian language and *es_core_news_lg* for Spanish contain 500 thousand unique vectors based on written news and media text. The spaCy model (figure 9) was used to predict negative or positive sentiment over untrained comments by including the *textcat* component, which serves for document classification.

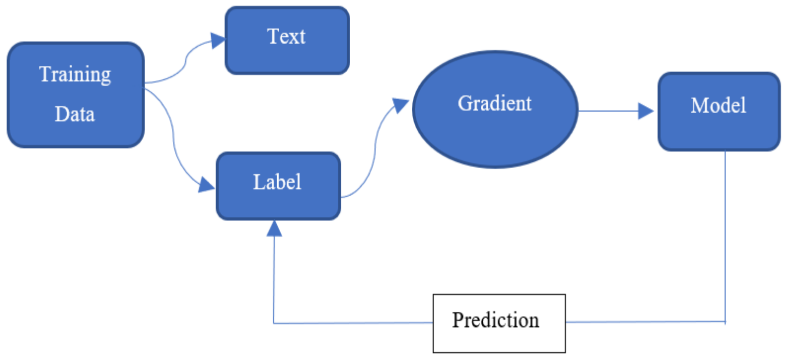

Figure 9 - Structure of classification algorithm

Since the language expressed in comments of both newspapers can be compared to the language posted in Twitter posts, both algorithms were trained with training data obtained from Twitter. The normalised pre-labelled text data was split into a training and a validation set. The ratio was to split the data into 75 % training data and 25% test data. The pre-classified training data was saved into two folders containing the negative and positive comments. This structure was used to create the “positive” and “negative” labels of the pipeline in a dictionary structure. The text and its corresponding label were then saved into a list. In order to remove the bias caused by the order of the training data, the library *random* was used to shuffle the data using the *shuffle ( )* function.

After loading the pretrained spaCy pipeline, the *textcat* component was created with the *.create_pipe ( )* function and then added to the pipeline using the *.addpipe ( )* function to train the model. The “negative’ and “positive” labels were then added to the *textcat* pipe. The training was then performed using a for loop with specified iterations.

After saving the training model the unclassified comments were fed into the algorithm to return a positive and negative output.

General Sentiment Classification

	Precision	*Recall*	*F-Score*
Italian model	0.81	0.81	0.81
Spanish model	0.87	0.85	0.87

Table 4 - Model evaluation

Using the training data obtained from Twitter sentiment data sets, the model showed a good performance in both cases. While the Italian model showed an F-Score of 0.81, the Spanish model performed slightly better with a F-Measure of 0.87 (see table 5).

6.2.1 Applying the algorithm to COVID-19 comments

The output of the algorithm was saved in a separate python list, containing the sentiment score per month of both corpora. The sentiment analysis showed a great level of negativity throughout the pandemic for both countries. The following plots will show the analysis results of both countries in terms of the determination of the general sentiment and application of context to enhance the analysis results.

6.2.2 Analysis of general sentiment in Spain

The analysis of the containment measures in chapter three showed that the Spanish government had a more passive approach, when the pandemic was handled. In comparison to Italy, the measures in Spain were applied at a later stage as the government tried to postpone a nationwide lockdown as much as possible. As expected, negativity was the dominating sentiment during the pandemic. As shown in figure 10, the algorithm classified the months January, March, May, October, November and December as highly negative, while the intensity shifted in other months.

General Sentiment Classification

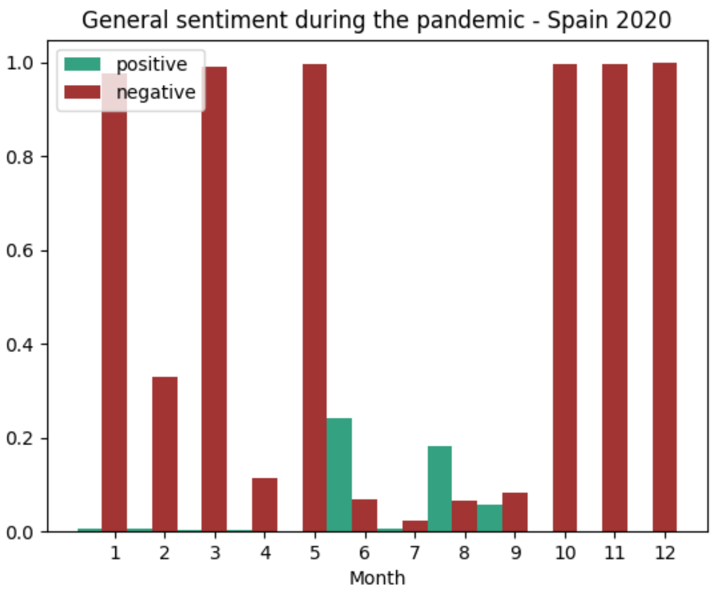

Figure 10 - General sentiment in El Mundo comments

The negativity did however remain at high rate throughout the whole year with different intensity depending on the period. While the negative sentiment always remained at a high level, a spike of positivity could be detected in the period from the end of April to the end of August (see figure 11).

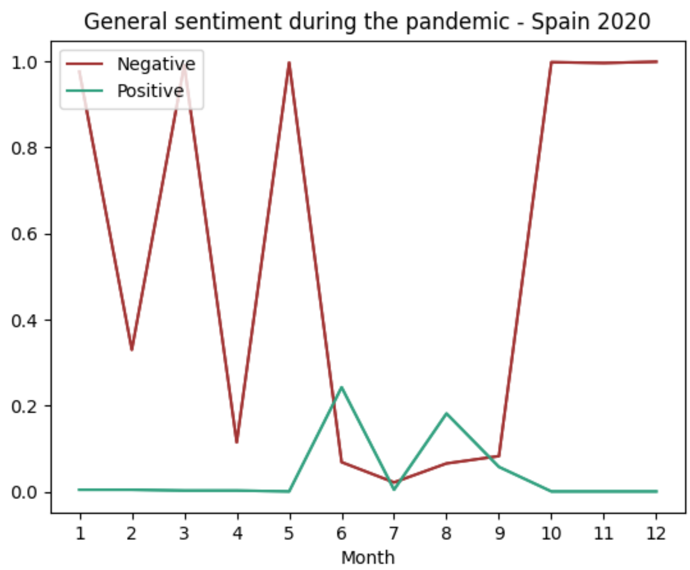

Figure 11 - General sentiment in El Mundo comments (graph)

This can be explained with the dropping numbers of cases during the summer of 2020, which applied to the European continent and the resulting easement of restriction measures. Other changes and spikes can be put in relation to the applied government measures during the different waves and easement plans.

The graphical analysis in figure 12 shows the sentiment during different containment measures applied by the Spanish government.

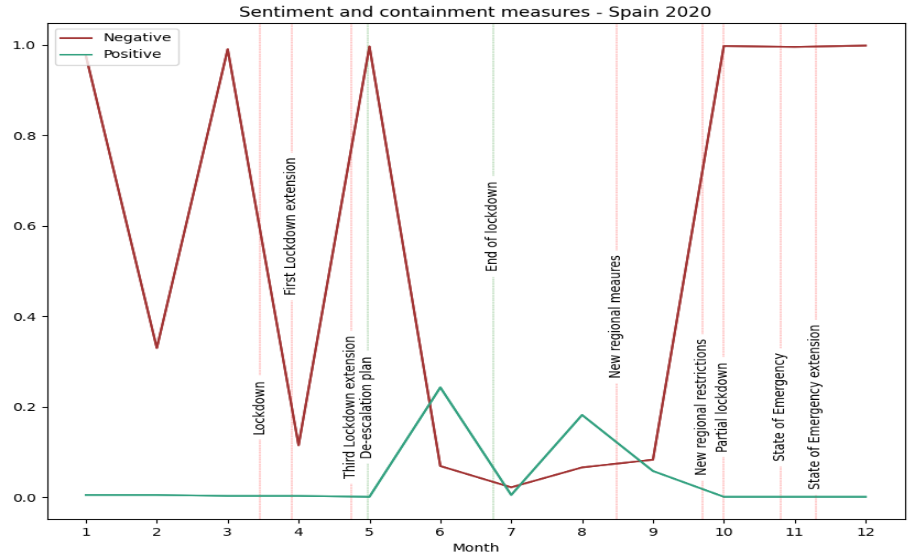

Figure 12 - General sentiment and containment measures - El Mundo comments

You can see a rising negativity during the first wave of the pandemic in March 2020. The readers appeared to comply with the imposed lockdown at first but were negatively impacted by the lockdown extension. The sentiment shifted to a more positive note when the de-escalation plan was announced in the beginning of May. A highly positive sentiment could be detected when the easement measures were applied, which can be seen in two spikes of positivity in the green curve. Nonetheless, the level of negativity rose to new heights with the announcement of new measures, which first applied for a regional basis, but soon later applied for the whole country. Therefore, from September onwards the negativity remained at a very high level until the end of the year. The new rise in cases and the measures applied as consequence spiked negativity in the population during the second wave of the pandemic.

6.2.3 Analysis of general sentiment in Italy

The analysis of Italian comments in La Repubblica showed a more negative picture, as seen in figures 13 and 14. High levels of negativity were detected from January to May 2020 and in July, August, and November. The general sentiment showed a higher volatility in comparison to the general sentiment in comments from El Mundo.

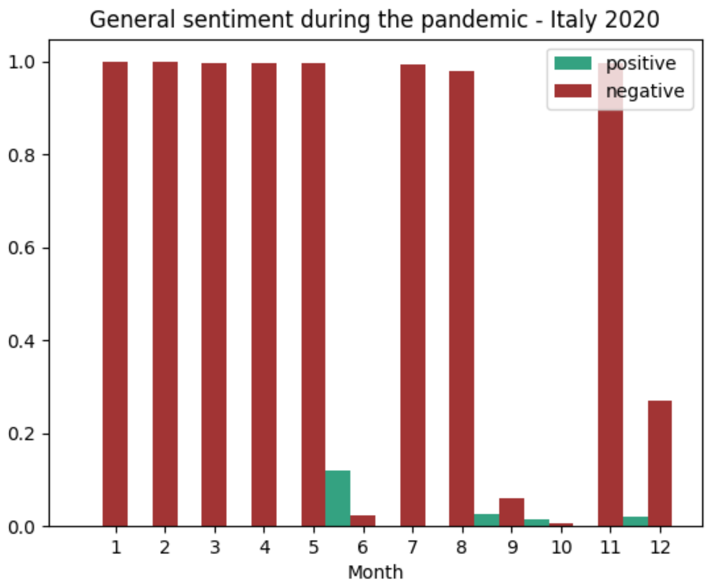

Figure 13 - General sentiment in La Repubblica comments

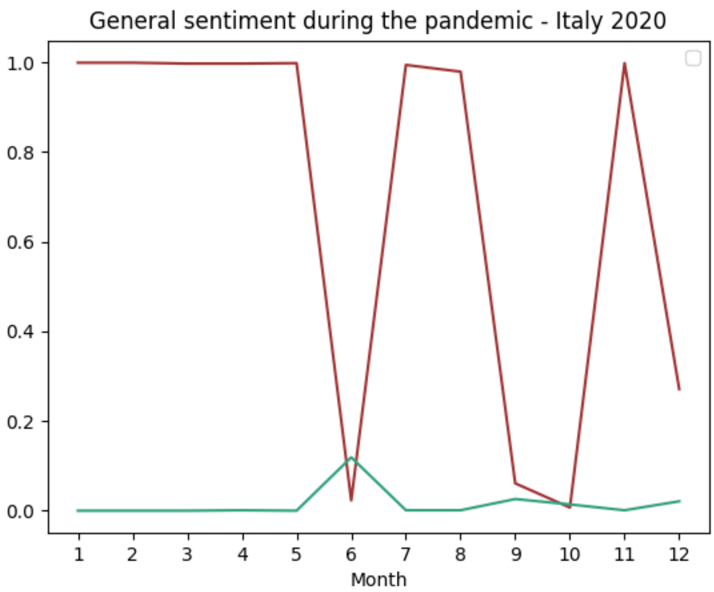

Figure 14 - General sentiment in La Repubblica comments (graph)

The sentiment analysis showed its highest positivity in June, which can also be explained with the dropping numbers and the easement of containment measures as seen in figure 15.

General Sentiment Classification

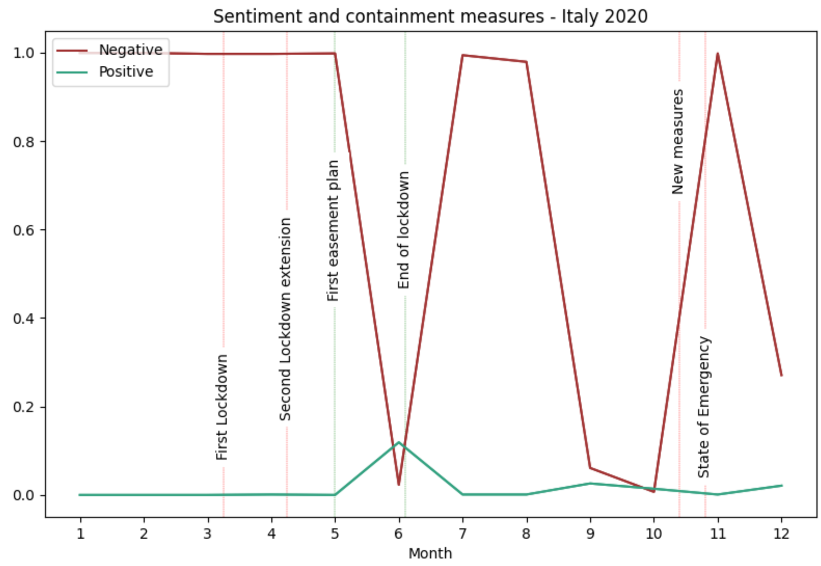

Figure 15 - Sentiment and containment measures - La Repubblica comments

The negativity rose after the end of the nationwide lockdown and remained high until August. With the new measures and the reinstatement of the state of emergency in October, the negativity began to rise again in October.

At the same time, a rise of negativity was also detected on comments from July and August, although the lockdown was ended, and the numbers kept decreasing. An analysis of the comment section in July and August showed the topics that dominated during the phase of negativity. The word clouds in figures 16 and 17 show the key points of discussion in the comment section during both months.

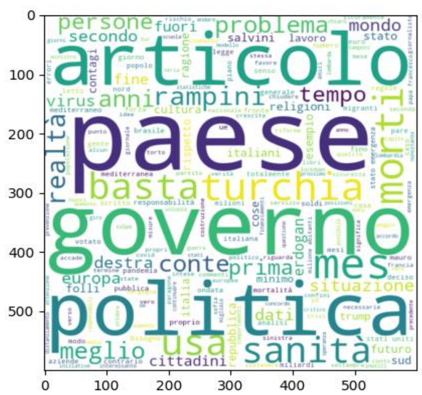

Figure 16 - Comments of La Repubblica in July (Word cloud)

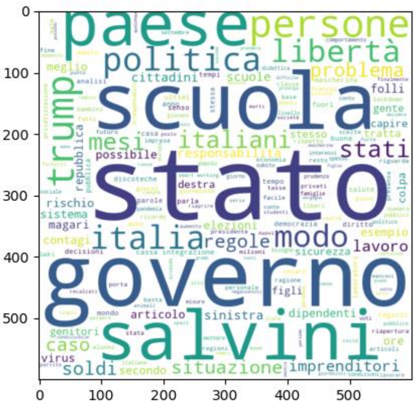

Figure 17 - Comments of La Repubblica in August (Word cloud)

In July the most common topics were politics, the government, the Italian nation and the article itself. Other government entities, such as Prime Minister Conte and former minister of interior Matteo Salvini were also frequently mentioned, as well as Europe and the European Stability Mechanism (it: MES). In August most discussed topics were also related to politics, with the state and country being frequently mentioned, as well the government and Salvini. One of the main topics were schools and education as well. The Italian government opened schools again in August, despite the number of cases rapidly increasing in the time period. The wearing of masks inside schools was also controversially discussed. Along with the rising number and the general concern of a second COVID-19 wave and the reinstatement of free movement which some considered to be too soon this could explain the negativity spike in July and August. Whether the negativity was directly directed at the government and government measures, will be explored in the next chapter by performing an aspect-based emotion analysis.

7 Emotion Detection

In order to determine different basic emotions expressed in the comment section during the pandemic, a rule-based key word recognition method was used using the National Research Council of Canada (NRC) lexicon. Jockers' Syuzhet R Package was then used to analyse the comments using sentence vectorisation and emotion scores calculated with NRC and saved in an output file. The output file was used to analyse the level of trust in comments mentioning the government and government entities. Furthermore, the levels of fear and trust were analysed in the context of the first nationwide lockdown imposed by the Italian and Spanish governments. Finally, an analysis was made to determine if the scores of opposite emotions returned a negative correlation, as postulated in Plutchik's emotion theory.

7.1 NRC Word-Emotion Association Lexicon

The NRC Word-Emotion Association Lexicon, formerly known as *EmoLex*, was created by Peter Turney and Saif M. Mohammad in 2010. Carefully chosen target words, that include most common nouns, verbs, adjectives and adverbs, as well as most common bigrams, were annotated with an emotion value (Mohammad et. al. 2010, 27). NRC uses Plutchik's concept of eight basic emotions of joy, sadness, anger, fear, disgust, surprise, anticipation and trust. Using crowdsourcing a set of words was annotated manually by English speakers. In order to maintain a high annotation quality, malicious entries were filtered out using a questionnaire, which determined whether the annotation task was performed in a serious manner and if the annotator had a fluent or native understanding of the English language (Mohammad et. al. 2010, 28f.).

The annotators were given example words related to different emotions and annotated the terms intuitively. If a term provoked and could be associated with different emotions, it was annotated accordingly. The final annotation was determined with agreement values, which determined the percentage of times the majority class agreed with the annotation (Mohammad et. al. 2010, 31).

With this approach the target words were annotated according to their sense and not at word level. In total, 14183 terms were annotated and translated to over a hundred of other languages (Mohammad et. al. 2013, 31). The NRC emotion lexicon can be downloaded separately to perform emotion analysis with any desired programming language.

7.2 Syuzhet Package

The Syuzhet package was created by Matthew L. Jockers to determine the relation of sentiment and plot information in fiction. With the Syuzhet dictionary, bing, afinn and nrc, Jocker's Syuzhet package integrates four different dictionaries and provides methods and functions to extract sentiment and emotion data from text. The Syuzhet lexicon is used as default and contains over 10000 words with an associated sentiment value, ranging from -1 to 1. The bing lexicon, developed by Hu et al. is made of 6780 words. The afinn lexicon, developed by Nielsen includes obscene and slang words from the Internet and is gradually extended using the social media platform Twitter. This study will use the NRC dictionary, since it integrated both Spanish and English emotion words and phrases.

The input text is loaded, parsed and converted into a vector of sentences using its *get_sentences ( )* function. The returned vector contains one item per tokenised sentence. This study used the NRC method to determine the sentiment values using the *get_nrc_sentiment ( )* function. The vector of sentences was fed into the function to create an NRC vector containing the values of each sentence and each determined sentiment value. The function finds all lexicon words contained in each sentence and computes the sum of all sentiment values. If a word occurs in one category the value 1 is added. Each sentence gets a score for each sentiment category (Naldi 2019, 4f.). In this study, the results were also transposed into a data frame plotted to determine the dominating emotions.

7.3 General emotions expressed during the pandemic

In order to analyse the emotion, the comment data set was split by month. The Syuzhet R package was then used to firstly determine the score of each emotion detected in every comment and secondly, to create a score table for each month containing each emotion score detected by the algorithm.

Emotion Detection

	sentiment	count
1	anger	181324
2	anticipation	168346
3	disgust	147315
4	fear	289198
5	joy	126391
6	sadness	259494
7	surprise	97527
8	trust	278216

Table 5 - Emotion count results - El Mundo

The result tables (see table 6) were then used to calculate the average emotion score contained in each month. The results in figure 18 show that trust, fear and sadness constantly were the dominant emotions in the El Mundo comment section. Positive and neutral emotions such as anticipation, joy and surprise remained at a low rate throughout the analysed period.

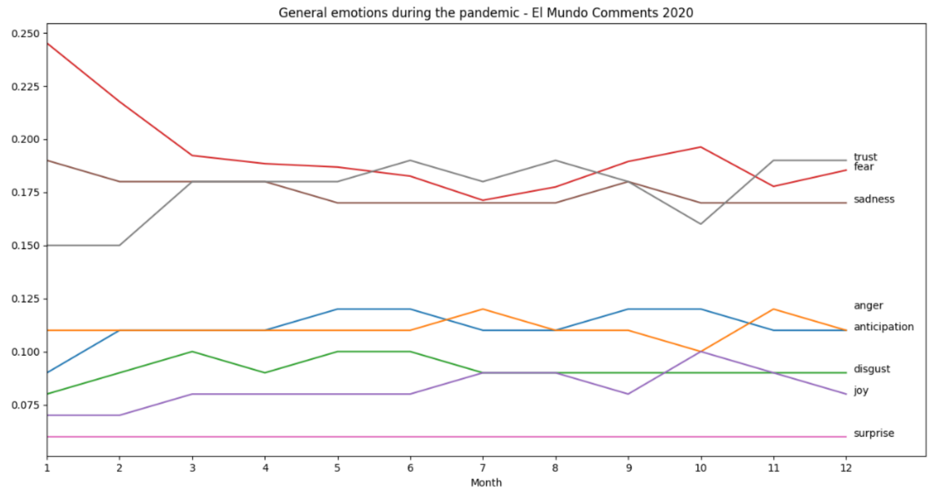

Figure 18 - General emotions - El Mundo comments

The Italian data in figure 19 shows a similar picture, with trust being the most dominant emotion, followed by fear and sadness. The level of trust, however, remained the most dominant emotion throughout the whole period. Other positive and neutral emotions such as anticipation, joy and surprise remained at a low rate as well.

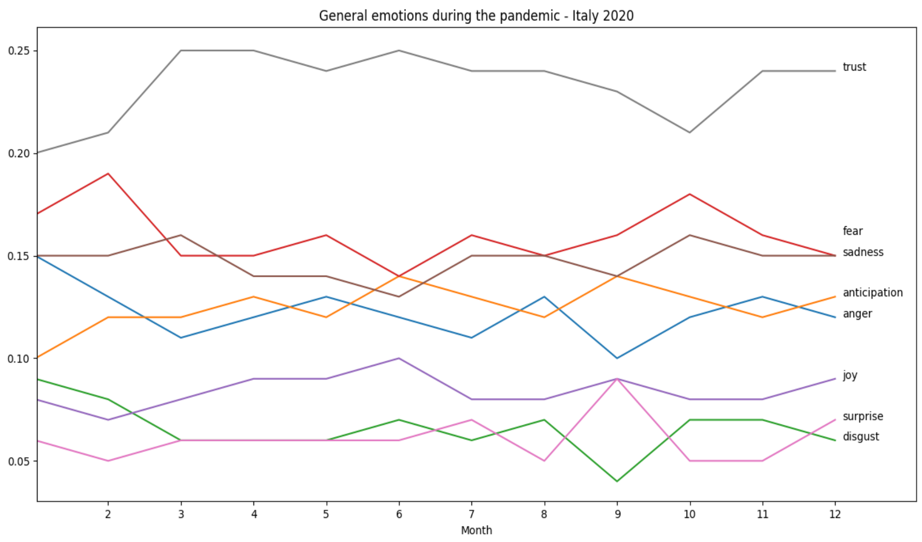

Figure 19 - General emotions - La Repubblica comments

The emotion results confirm the dominant negativity detected with the sentiment analysis algorithm in chapter 6. Table 7 shows that both comment sections had the same dominating emotions throughout the whole analysed period, with fear, sadness and trust all shifting their ranks throughout different periods in El Mundo comments and fear and sadness constantly shifting their ranks in the Italian comment section. Trust was the constant dominant emotion in the La Repubblica dataset.

Month	El Mundo	La Repubblica
January	1.FEAR 2.SADNESS 3.TRUST	1.TRUST 2.FEAR 3.SADNESS
February	1.FEAR 2.SADNESS 3.TRUST	1.TRUST 2.FEAR 3.SADNESS
March	1.FEAR 2.SADNESS 3.TRUST	1.TRUST 2.SADNESS 3.FEAR
April	1.FEAR 2.TRUST 3.SADNESS	1.TRUST 2.FEAR 3.SADNESS
May	1.FEAR 2.TRUST 3.SADNESS	1.TRUST 2.FEAR 3.SADNESS
June	1.TRUST 2.FEAR	1.TRUST 2.FEAR

	3.SADNESS	3.SADNESS
July	1.TRUST 2.FEAR 3.SADNESS	1.TRUST 2.FEAR 3.SADNESS
August	1.TRUST 2.FEAR 3.SADNESS	1.TRUST 2.SADNESS 3.FEAR
September	1.FEAR 2.SADNESS 3.TRUST	1.TRUST 2.FEAR 3.SADNESS
October	1.FEAR 2.SADNESS 3.TRUST	1.TRUST 2.FEAR 3.SADNESS
November	1.TRUST 2.FEAR 3.SADNESS	1.TRUST 2.FEAR 3.SADNESS
December	1.TRUST 2.FEAR 3.SADNESS	1.TRUST 2.FEAR 3.SADNESS

Table 6 - Dominant emotions

Since the most dominant emotions throughout the whole analysed period were fear and trust, the next chapters will perform a more detailed analysis based on them. Firstly, the changes during the first nationwide lockdown will be analysed. Secondly, it will be determined if the opposites of the scores of these emotions show a negative correlation to their polar opposites to determine whether Plutchik’s theory applies to this context. Finally, the level of trust expressed in comments mentioning the government will be analysed as well.

7.4 Measuring fear and trust during first nationwide lockdown

To analyse the emotions detected in the comment section during the lockdown period, the dataset was split into the period of lockdown form March 14 to May 04 in Spain and from March 09 to May 04 in Italy. The general analysis for the emotion expressed during the whole lockdown period determined fear, sadness and trust as the dominant emotions (see figure 20 and 21). Fear dominated the comment section of El Mundo, while a very high level of trust was found in the comments of La Repubblica during lockdown.

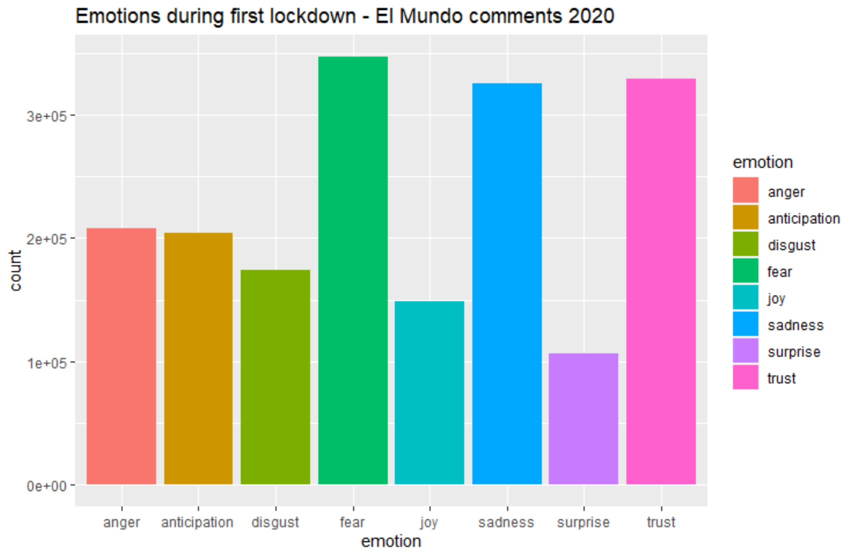

Figure 20 - Emotions during first lockdown - El Mundo comments

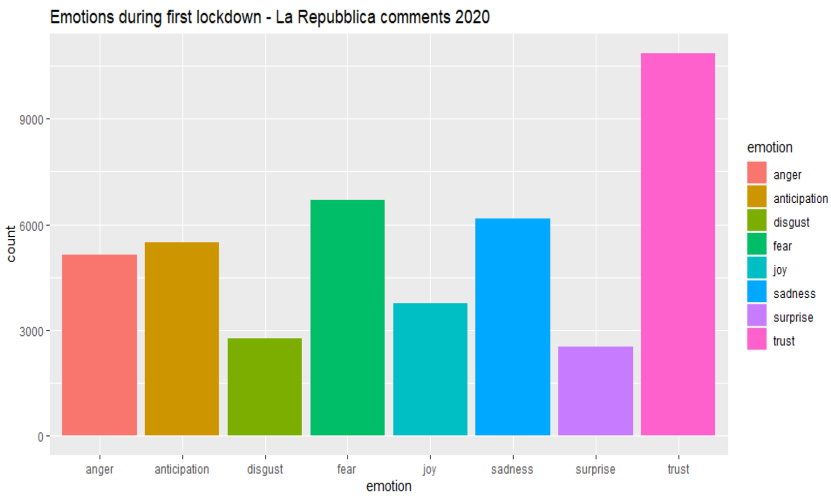

Figure 21 - Emotions during first lockdown - La Repubblica comments

Since the scores showed statistical outliers, the moving average of the scores was determined to minimise any bias caused by them.

The comment section of El Mundo showed similar reactions to each lockdown extension. The level of trust began to increase after the announcement of the first nationwide lockdown but decreased with the beginning of the first lockdown extension. This course could be observed within the first, second and third lockdown extension as well. After or shortly before each announcement of an extension, the trust decreased, but began to recover and even increase towards the end before it plummeted again after a new extension announcement (see figure 22).

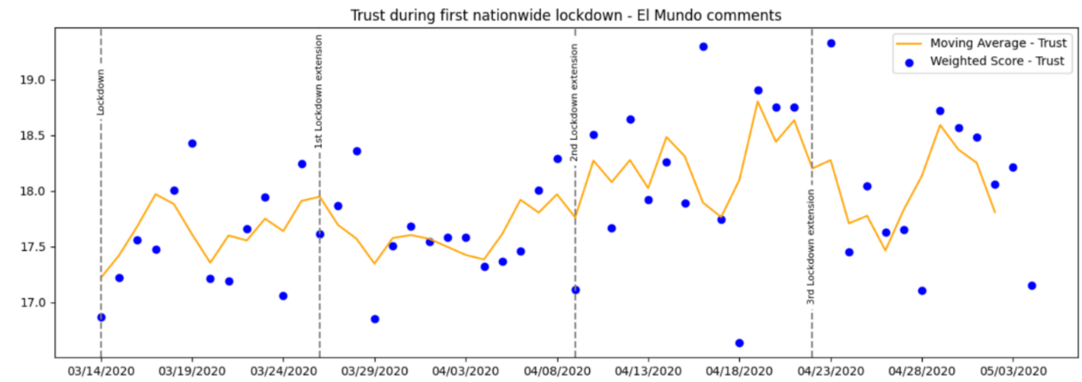

Figure 22 - Trust during first lockdown - El Mundo comments

A different picture was found in the Italian comment section with the level of trust plummeting during the first nationwide lockdown but increasing after each extension announcement. The level of trust remained at a generally higher level, compared to El Mundo comments (see figure 23).

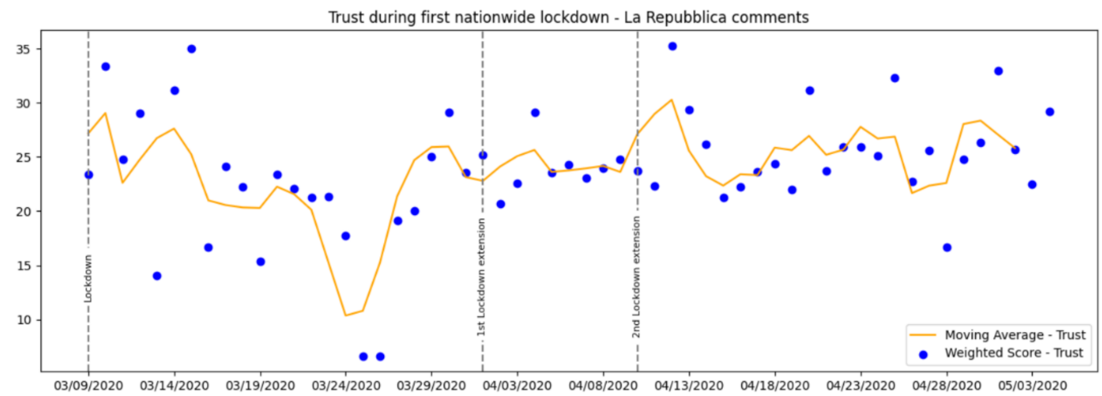

Figure 23 - Trust during first lockdown - La Repubblica comments

The analysis of the fear in the comment section of El Mundo (see figure 24) showed an initial decrease of fear during the first days of the nationwide lockdown, which increased in later days and significantly decreased towards the end.

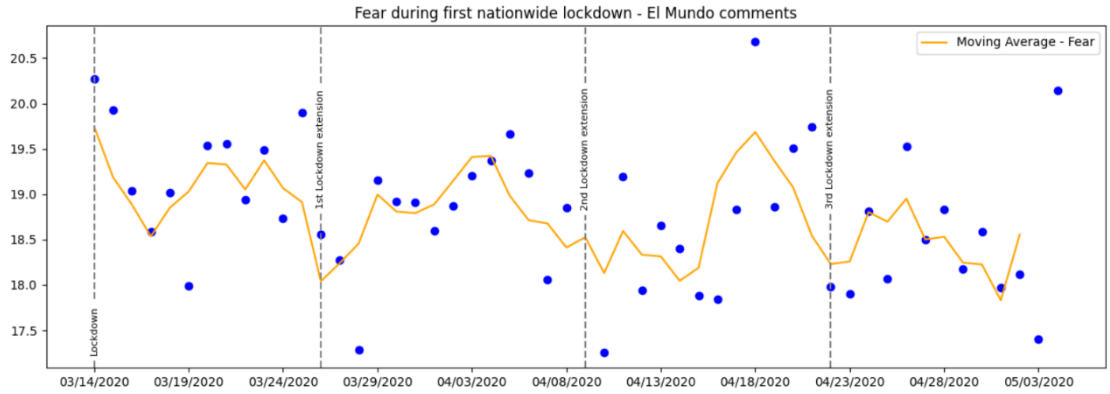

Figure 24 - Fear during first nationwide lockdown - El Mundo comments

After each lockdown extension the level of fear began to rise again, while plummeting towards the end. This could be observed after each lockdown extension. In the end of the 3rd extension the numbers plummeted and reached its lowest point, indicating the end of the lockdown phase along with the decreasing COVID-19 cases.

The graph of La Repubblica (figure 25) shows an increase of fear during the first phase of the nationwide lockdown, which reached its peak in the last week of March.

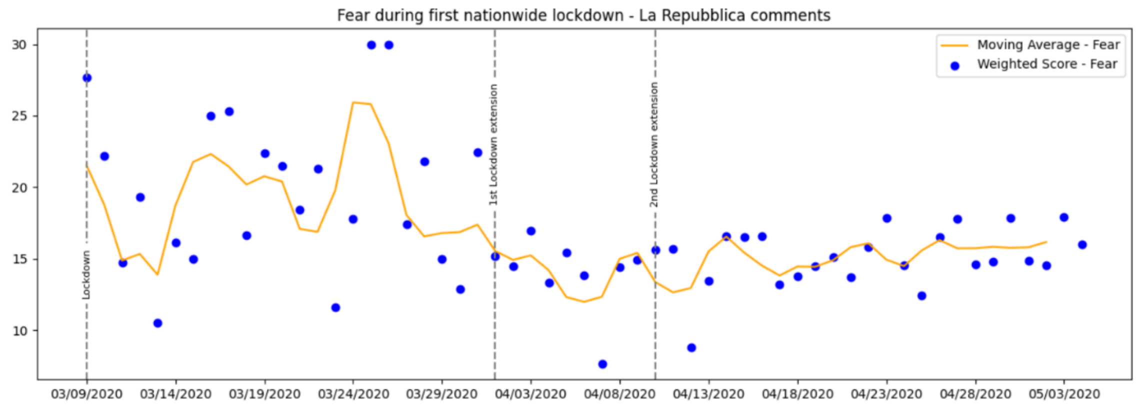

Figure 25 - Fear during first nationwide lockdown - La Repubblica comments

Towards the end of March and with the announcement of the first and second lockdown extension the level of fear noticeably decreased and remained at a low rate until the end of the nationwide lockdown.

7.5 Testing Plutchik's theory of emotional opposites

Robert Plutchik postulates that each primary emotion can be paired with its polar opposite which is represented in the wheel of emotion. Taking this into consideration, the dominant emotions fear and trust were compared and analysed with its opposite pairs anger and disgust.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Where,

r = Pearson Correlation Coefficient

x_i = x variable samples $\quad$ y_i = y variable sample

$\bar{x}$ = mean of values in x variable $\quad$ $\bar{y}$ =mean of values in y variable

Figure 26 - Pearson coefficient

This was tested by calculating the correlation of the scores of the opposite emotions. In general terms, correlation determines whether there is a monotonic relationship between two variables. In correlated data the change in one magnitude should be associated with a change of another magnitude. This can be applied to both directions. Values with a positive correlation mean that higher values of one variable cause higher values of another variable. Inversely, this means that a negative correlation describes the relationship in which a higher value of one variable, returns a lower value of another variable.

The relation of a change of one variable causing another variable to change as well is described as the covariance. The Pearson coefficient *r* (figure 26) is used to measure the covariance and returns a range between -1 and 1 (cf. Schober et al. 2018, 1763).

Applied to the emotion scores found in the comment sections this would mean that there should be negative correlation between the values found with the opposite emotions (e.g., trust and disgust), if Plutchik's theory applies. The analysis of trust and disgust are shown in figure 27.

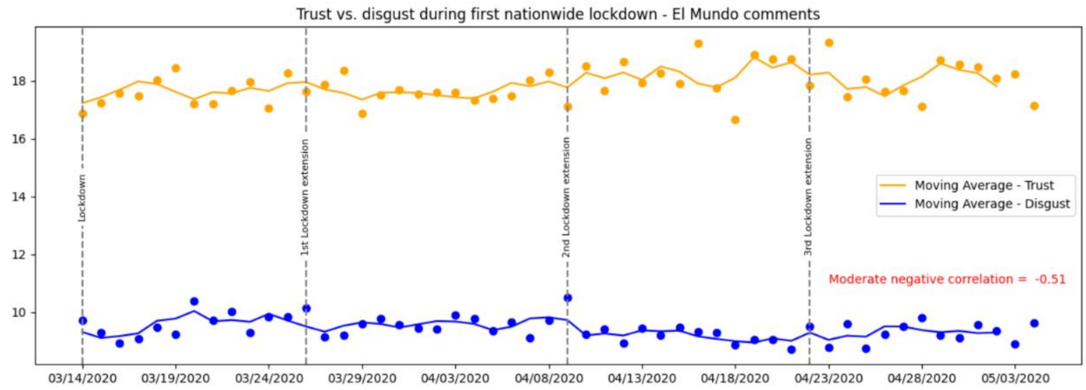

Figure 27 - Trust vs. disgust during first nationwide lockdown – El Mundo comments

It shows that the level of trust was generally significantly higher than the level of disgust. The calculation of the Pearson coefficient determined a moderate negative correlation of -0.51 between the two emotion scores.

The comments of la Repubblica showed an even higher trust rate and a lower disgust rate than El Mundo comments. The calculation of the Pearson coefficient also returned a moderate negative correlation of -0.57 (figure 28).

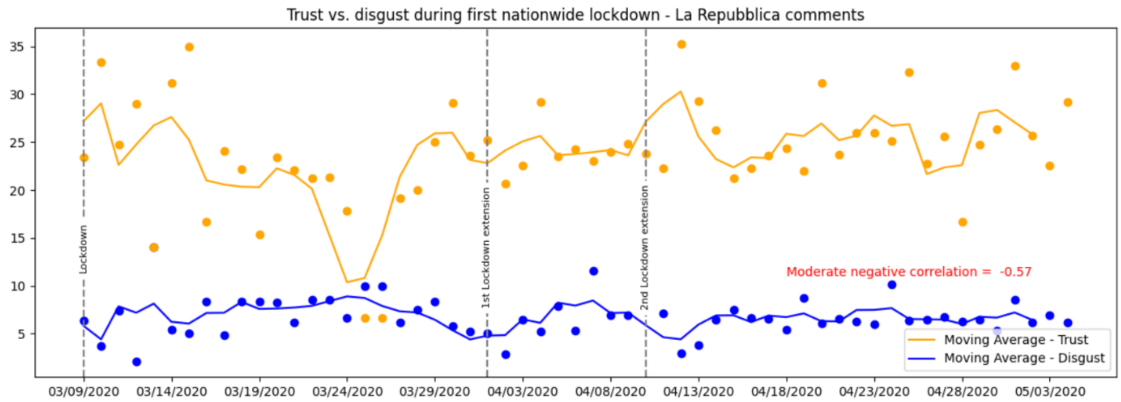

Figure 28 - Trust vs. disgust during first nationwide lockdown - La Repubblica comments

El Mundo comments also showed a high level of fear between 18 and 20 %, with a relatively low level of anger of ca. 11 %. The Pearson coefficient also returned a moderate negative correlation of -0.57 in this case (figure 29).

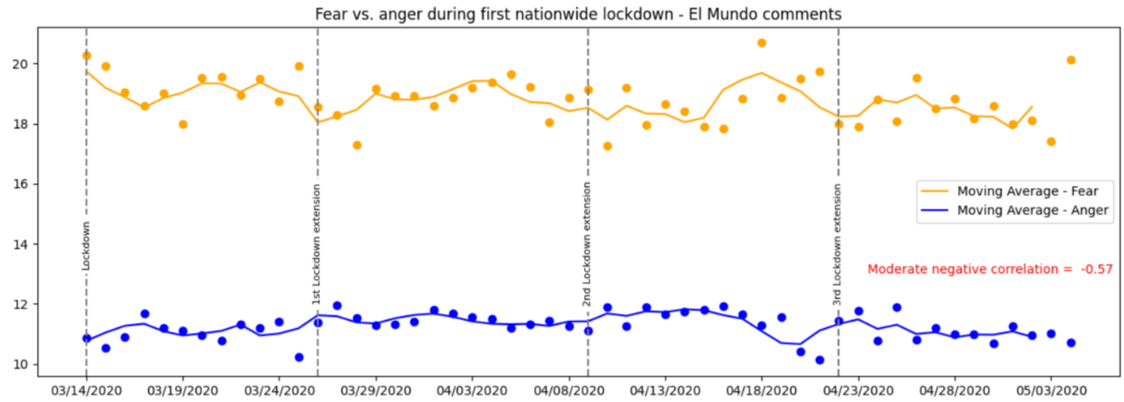

Figure 29 - Fear vs. anger during first nationwide lockdown - El Mundo comments

The comments of La Repubblica showed a higher level of anger, when compared to the comments of El Mundo. The level of fear remained higher than the level of anger. The datasets showed a low negative correlation of -0.29 (figure 30).

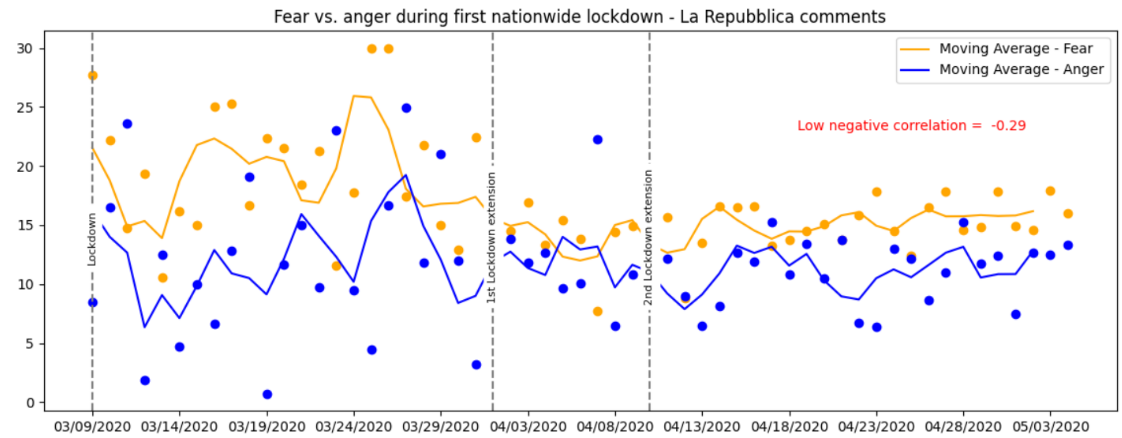

Figure 30 - Fear vs. anger during first nationwide lockdown - La Repubblica comments

Overall, the results show an existing negative correlation between the analysed dominant emotions and their opposites. This means that Plutchik's theory can certainly be applied to the analysed comment sections. During the lockdown period within this dataset, the level of disgust generally dropped whenever the level of trust rose, while the level of anger dropped, whenever the level of fear rose, and vice versa.

7.6 Measuring government trust

In order to analyse the government comments a filter was applied to the dataset, contained tokens about the government, for example *M5S (Movimento Cinque Stelle)*, *Conte* and *ministerio* for Italian comments or *gobierno*, *Sánchez and PSOE* for the Spanish dataset. The algorithm determined similar ranks of emotion for El Mundo comments mentioning the government. Fear dominated the comment section of El Mundo with trust having the second highest rate (see figure 31).

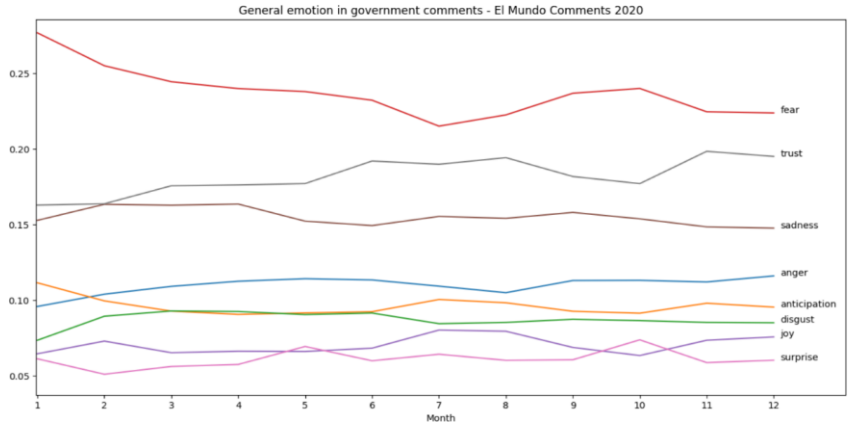

Figure 31 - General emotion - Comments about government - El Mundo

Nonetheless, the La Repubblica dataset showed a significant decrease of trust, with fear becoming more dominant in comments mentioning the government (see figure 32).

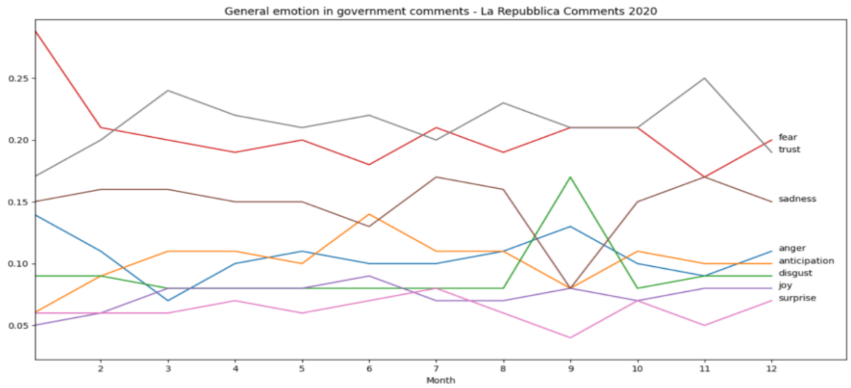

Figure 32 - General emotion - Comments about government - La Repubblica

Furthermore, the analysis of La Repubblica comments showed an increase of fear and a drop of trust levels after the end of the lockdown period and the beginning of restriction easements. The trust increased again after the announcement of new measures in October 2020, while fear dropped at the same time. A low negative correlation of -0.4 could indicate that trust decreased as fear increased (see figure 33).

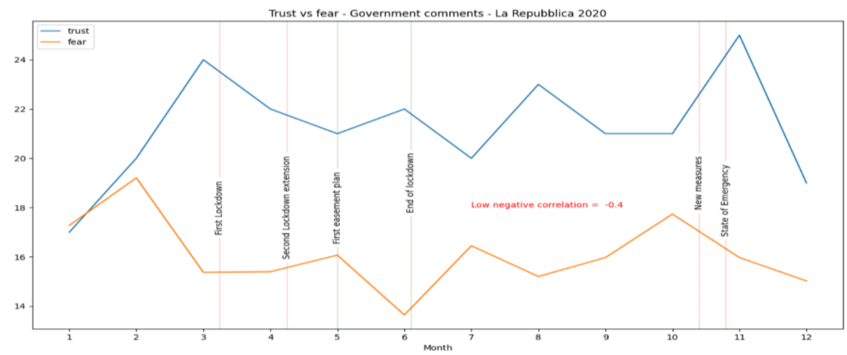

Figure 33 - Trust and fear in comments mentioning government - La Repubblica

El Mundo comments showed a clearer picture with a very high negative correlation of -0.9 detected between fear and trust scores. After the announcement of the first lockdown fear slightly dropped as trust increased. The fear rate dropped to its lowest point after the announcement of the end of lockdown, as the level of trust reached its peak shortly after. However, fear increased again with rising COVID-19 cases from August onwards, causing the trust to drop significantly. After the announcement of the partial lockdown in October fear decreased again, as the level of trust rose (see figure 34).

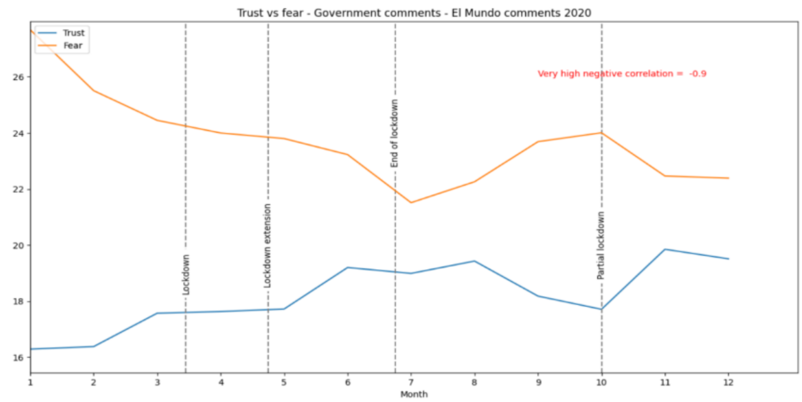

Figure 34 - Trust and fear in comments mentioning government - El Mundo

8 Conclusion

The aim of this study was to determine whether there was an existing relationship between shifting changes of sentiment and emotion and the application of containment measures during the pandemic in 2020, as well as to determine whether there was a relationship between opposite emotions expressed within this context. The sentiment analysis was performed using an algorithm created with Python’s spaCy library and uses a convolutional neural network to classify negative and positive comments, using pre-labelled training data obtained from Twitter. The Italian model achieved an F-Score of 0.81 and an F-Score of 0.87 was measured for the Spanish model.

The algorithm determined a great level of negativity throughout the analysed period from January 2020 to December 2020. In both cases a spike of positivity was detected during the summer with daily COVID-19 cases dropping and both governments thus easing the applied restrictions. The negativity increased again from the beginning of the second wave onwards. Applying the context of the restriction measures with a graphical analysis, the results showed an existing relationship between the application of the first nationwide lockdown and its according extensions. In El Mundo comments, each announcement of new measures and lockdown extensions caused a spike of negativity. The same applies to La Repubblica except for comments made in July and August, which showed a new negative spike. A separate analysis showed that the main topics of the negative comments made during these months were related to politics and applied measures.

The sentiment analysis was further extended using a dictionary-based emotion detection based on the NRC Lexicon and the Syuzhet Package. The categorisation of emotions was based on Robert Plutchik’s Psychoevolutionary theory of emotion. In both newspapers trust, fear and sadness dominated the comment section. Apart from trust, positive emotions, such as joy, anticipation and surprise remained at a low rate throughout the whole period in both newspapers. Furthermore, a separate day-by-day analysis was made for comments during the first nationwide lockdown of each country. During this time frame trust and fear dominated the comments. The context analysis showed that the level of trust in El Mundo increased after each announcement of lockdown, but constantly plummeted towards the end and with the announcement of new lockdown extensions.

This didn't apply to La Repubblica comments, in which the trust increased after each extension announcement and generally remained at a high level.

The analysis of the fear levels showed that fear began to rise again after each lockdown announcement in El Mundo comments and plummeted towards the end, reaching its lowest point towards the end of the nationwide lockdown. La Repubblica comments showed a high level of fear during the first lockdown phase, which reached its highest point during the last week of March and noticeably decreased after the first and second lockdown extension. The level of fear remained low until the end of the nationwide lockdown. Furthermore, the relationship between the dominant emotions fear and trust, and their polar opposites, according to Plutchik's PTE model was analysed using the Pearson correlation coefficient. The analysis determined a moderate negative relationship between the pairs trust and disgust and fear and anger. This showed that Plutchik's theory could be applied to the emotions expressed in the analysed comment section. Finally, the level of trust was analysed in comments mentioning the government and government entities. A moderate level of correlation was detected between fear and trust in La Repubblica comments, and a very high level of correlation was detected in El Mundo comments. This showed that the level of trust dropped, whenever the level of fear increased.

The results [13]show that the emotional reaction related to the government applied measures can be tracked using emotion detection and sentiment analysis. The insights gained from this work can thus be used to trace the reaction of the population to general political decisions. Governments and other decision-making institutes can use the same approach in their handling of future disasters and catastrophes by considering the emotional reaction of the population. Furthermore, future related studies can use a larger amount of data by including new data obtained from 2021 and/or other newspapers. This could help with other political decisions, for example by tracing the reactions to COVID-19 vaccines or other treatments. Furthermore, using the press corpora, the analysis of sentiment can be further enhanced by including the aspect of multimodality in the form of images, videos and audio. Moreover, the press corpora created within this work can be used for other linguistic, sociolinguistic, and neuroscientific studies. Furthermore, the press corpora are part of the larger project *Linguistic change and multimodal communication*

[13] All source and programme codes created to obtain these results have been added to the attached CDs.

in times of the COVID19 pandemic, headed by PD Dr. Anna Ladilova in the Department of Romance Studies at JLU, which analyses language contact and the formation of new words in the context of the pandemic in Spanish, Portuguese, Italian and French newspapers. In this context, the press corpora will be extended and made available to other researchers.

9 Bibliography

- Ananasso, A., Da Stabile, E., Gagliardi, G., Matteucci, P. (2020): "Coronavirus in Italia: tutte le notizie di febbraio" (https://www.repubblica.it/cronaca/2020/02/22/news/coronavirus_in_italia_aggiornamento_ora_per_ora-249241616/ -- last accessed: 12.11.2021).
- Aslam, M., Awan, T., Syed, J. H., Kashif, A., Parveen, M. (2020): "Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak", in: *Humanities and Social Sciences Communications.* 7, 23, n.pag.
- Barone, N., Gagliardi, A. (2020): "Dal 3 giugno l'Italia riapre le frontiere con l'Ue, via libera a spostamento tra regioni" (https://www.ilsole24ore.com/art/stop-ponte-2-giugno-ecco-cosa-prevede-schema-dl-riaperture-ADWEtlQ -- last accessed: 12.11.2021).
- BBC (2020): "Coronavirus: Italy extends emergency measures nationwide" (https://www.bbc.com/news/world-europe-51810673 – last accessed: 12.11.2021).
- BBC (2020): "Madrid coronavirus: Spain orders lockdown amid rise in cases" (https://www.bbc.com/news/world-europe-54364015 – last accessed: 12.11.2021).
- Biarella, L. (2020): "Coronavirus: lockdown fino al 3 maggio" (https://www.altalex.com/documents/news/2020/04/14/coronavirus-lockdown-fino-al-3-maggio – last accessed: 12.11.2021).
- Castro, I. (2020): "Pedro Sánchez prorrogará el estado de alarma hasta el 26 de abril pero irá abriendo la actividad económica" (https://www.eldiario.es/politica/pedro-sanchez-prorrogara-alarma-proximo_1_1213467.html -- last accessed: 12.11.2021).
- Ciriaco T, Rubino, M., Ziniti, A. (2020): "Coronavirus, Fase 2: dal 4 maggio sì a incontri con familiari. Il 18 riapriranno i negozi, il primo giugno bar, ristoranti, parrucchieri e centri estetici" (https://www.repubblica.it/politica/2020/04/26/news/coronavirus_riaperture_cabina_regia_governo_regioni-254928829/ -- last accessed: 12.11.2021).
- Costola, M., Nofer, M ; Hinz, Ol.; Pelizzon, L. (2020): "Machine learning sentiment analysis, COVID-19 news and stock market reactions", in: *SAFE working paper series 288*.
- Cottone, N. (2020): "Coronavirus: chiusura fino al 13 aprile. Stretta sullo sport. Conte: «Se allentiamo ora, sforzi vani»" (https://www.ilsole24ore.com/art/coronavirus-lockdown-fino-13-aprile-stretta-allenamenti-sportivi-professionisti-ADCwGVH?refresh_ce=1 – last accessed: 12.11.2021).
- Das, S., Dutta, A. (2020): "Characterizing public emotions and sentiments in COVID-19 environment: A case study of India", in: *Journal of Human Behavior in the Social Environment,* 31:1-4, 154-167.

- Departamento de Seguridad Nacional (DSN) (2020): “Coronavirus (COVID-19) - 29 de febrero 2020” (https://www.dsn.gob.es/es/actualidad/sala-prensa/coronavirus-covid-19-29-febrero-2020 -- last accessed: 12.11.2021).
- Departamento de Seguridad Nacional (DSN) (2020): “Coronavirus (COVID-19) - 31 de marzo 2020” (https://www.dsn.gob.es/es/actualidad/sala-prensa/coronavirus-covid-19-31-marzo-2020 -- last accessed: 12.11.2021).
- Ekman, P. (1992): “An Argument for Basic Emotions”, in: *Cognition and Emotion*, 169-200.
- Gazzetta Ufficiale della Reppublica Italiana (2020): “Decreto del presidente del consiglio dei ministri 4 marzo 2020” (https://www.gazzettaufficiale.it/eli/id/2020/03/04/20A01475/sg -- last accessed: 12.11.2021).
- Gazzetta Ufficiale della Reppublica Italiana (2020): “Decreto del presidente del consiglio dei ministri 24 ottobre 2020” (https://www.gazzettaufficiale.it/eli/id/2020/03/04/20A01475/sg -- last accessed: 12.11.2021).
- Gazzetta Ufficiale della Reppublica Italiana (2020): “Decreto del presidente del consiglio dei ministri 3 dicembre 2020” (https://www.gazzettaufficiale.it/eli/id/2020/12/03/20A06767/sg -- last accessed: 12.11.2021).
- Gazzetta Ufficiale della Reppublica Italiana (2020): “Decreto-Legge 7 ottobre 2020, n. 125” (https://www.trovanorme.salute.gov.it/norme/dettaglioAtto?id=76873 -- last accessed: 12.11.2021).
- Guerzoni, M. Sarzanini, F., Servegnini, C. (2020): “Decreto 18 maggio, il testo definitivo su riaperture e spostamenti firmato da Conte” (https://www.corriere.it/politica/20_maggio_17/decreto-18-maggio-testo-definitivo-riaperture-firmato-conte-a75f00be-985b-11ea-ba09-20ae073bed63.shtml -- last accessed: 12.11.2021).
- Honnibal, M., & Montani, I. (2017): *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*, to appear.
- Hernández, A. (2020): “Balance del Covid-19 en España durante el mes de agosto” (https://www.elmundo.es/espana/2020/08/30/5f494d2121efa0cb2a8b45bc.html -- last accessed: 12.11.2021).
- Imran, A., Daudpota, S., Kastrati, Z., Batra, R. (2020): “Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets”, in: *IEEE Access* 8, 181074 - 181090.
- Imbir, K.K. (2017): “Psychoevolutionary Theory of Emotion (Plutchik)”, In: *Zeigler-Hill V., Shackelford T. (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham. https://doi.org/10.1007/978-3-319-28099-8_547-1*.

Bibliography

- Izard, C. E. (1977). *Human Emotions*. New York: Plenum Press.
- Jockers, M. (2020): "Introduction to the Syuzhet Package" (https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html -- last accessed: 13.11.2021).
- Kaur, C., Sharma, A.(2020) : "Twitter Sentiment Analysis on Coronavirus using Textblob." (https://easychair.org/publications/preprint/Fd5m -- last accessed:12.11.2021).
- Kim, Y. (2014): "Convolutional Neural Networks for Sentence Classification", in: *arXiv:1408.5882v2 [cs.CL]*.
- Kolchyna, O., Souza, T.T.P., Treleaven, P.C., Aste, T. (2015): "Methodology for Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination.", in: *arXiv; Computation and Language*, n. pag.
- Kruspe A., Häberle M., Kuhn I., Zhu XX.(2020): "Cross-language sentiment analysis of European Twitter messages during the COVID-19 pandemic", in: *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*.
- Kumar, A., Khan, S. (2020): "COVID-19 pandemic: a sentiment analysis: A short review of the emotional effects produced by social media posts during this global crisis", in: *European Heart Journal*, Volume 41, Issue 39, 3782–3783."
- Lamet, Juanma (2020): "El plan del Gobierno para la Navidad: cenas de Nochebuena y Nochevieja de seis personas como máximo y toque de queda a la una" (https://www.elmundo.es/espana/2020/11/24/5fbcd33321efa066448b46ec.html -- last accessed: 12.11.2021).
- La Moncloa (2020): "El Gobierno aprueba prolongar el estado de alarma hasta el 10 de mayo" (https://www.lamoncloa.gob.es/consejodeministros/resumenes/paginas/2020/240420-consejo.aspx -- last accessed: 12.11.2021).
- La Repubblica (2020): "Coronavirus, terza vittima in Italia: è una donna ricoverata a Crema. 152 casi positivi" (https://www.repubblica.it/cronaca/2020/02/23/news/coronavirus_italia-249329434/ -- last accessed: 12.11.2021).
- Li, X., Zhou, M., Wu, J., Yuan A., Wu F., Li, J. (2020): "Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions", in: *arXiv:2005.14464*, n. pag.
- Linde, p. (2020): " Sanidad confirma en La Gomera el primer caso de coronavirus en España" (https://elpais.com/sociedad/2020/01/31/actualidad/1580509404_469734.html -- last accessed: 12.11.2021).
- Liu, B. (2nd edition 2020): „Sentiment Analysis. Mining Opinions, Sentiments, and Emotions ", in: Huang, C.: *Studies in Natural Language Processing. Cambridge*: Cambridge University Press, 1-432.

- Marcos, José (2020): ” España abrirá el 21 de junio sus fronteras con la UE salvo con Portugal, que será el 1 de julio” (https://elpais.com/espana/2020-06-14/espana-abrira-el-21-de-junio-sus-fronteras-con-la-ue-salvo-con-portugal-que-sera-el-1-de-julio.html -- last accessed: 12.11.2021).
- Mari, L., Rubino M. (2020): “Mascherine, obbligatorie sempre al chiuso tranne che a casa. Le multe restano invariate: da 400 a mille euro” (https://www.repubblica.it/politica/2020/10/07/news/mascherina_governo_cdm-269758198/ -- last accessed: 12.11.2021).
- Mayor Ortega (2020): “El Congreso aprueba con una mayoría aplastante prorrogar el estado de alarma hasta el 11 de abril” (https://www.lavanguardia.com/politica/20200326/4893390141/coronavirus-congreso-aprueba-prorroga-estado-alarma-aplastante-mayoria.html -- last accessed: 12.11.2021).
- McMurtry, A., Zampano, G. (2020): “COVID-19: What went wrong in Italy and Spain?” (https://www.aa.com.tr/en/europe/covid-19-what-went-wrong-in-italy-and-spain/1797461 -- last accessed: 12.11.2021).
- Ministero della Salute ´(2020): “Covid-19, in gazzetta ufficiale il Dpcm 1 marzo 2020. Tutte le misure di contenimento per zone” (https://www.salute.gov.it/portale/nuovocoronavirus/dettaglioNotizieNuovoCoronavirus.jsp?lingua=italiano&menu=notizie&p=dalministero&id=4137 -- last accessed: 12.11.2021).
- Ministerio de Sanidad (2020): “El Ministerio de Sanidad y las CCAA acuerdan por unanimidad actuaciones coordinadas para controlar la transmisión de la COVID-19” (https://www.mscbs.gob.es/gabinete/notasPrensa.do?metodo=detalle&id=5025 – last accessed: 12.11.2021)
- Mitchell, R. (2020): *Web Scraping with Python. Collecting data from the modern web*, Sebastopol: O’Reilly Media, Inc.
- Mohammad, S.M., Turney, P.D. (2013): “Crowdsourcing a Word–Emotion Association Lexicon”, in: *Computational Intelligence* 29.
- Munera, I. (2020): “Las claves del estado de alarma que ha aprobado el Gobierno: ¿qué medidas supone?” (https://www.elmundo.es/espana/2020/10/25/5f95484121efa0ee3d8b4628.html -- last accessed: 12.11.2021).
- Naldi, M. (2019): “A review of sentiment computation methods with R packages.”, in: *arXiv abs/1901.08319*, n. pag.
- Plutchik, R. (1980): Emotion. A Psychoevolutionary Synthesis., New York: Harper Row Publishers.
- Plutchik, R (1980): Theories of Emotion., London: Academic Press, Inc.
- Plutchik, R (1982): “A psychoevolutionary theory of emotions”, in: *Social Science Information* 21, 529-553.
- Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald B., Beltekian, D., Roser , M.(2020) :*Coronavirus*

Pandemic (COVID-19) – Published online at OurWorlInData.org.

- Roseman, I.J., Smith, C.A.(2001): “Appraisal Theory. Overview, Assumptions, Varieties, Controversies”, in: Scherer, K. R., Schorr, A., Johnstone, T.: *Appraisal Processes in Emotion: Theory, methods, research.*, New York: Oxford University Press, 3-19.
- Ruotolo, N. (2020): “More than 100,000 coronavirus cases reported in Italy” (https://edition.cnn.com/world/live-news/coronavirus-outbreak-03-30-20-intl-hnk/h_06feee39fa5d282b38f091cdce722abd -- last accessed: 12.11.2021).
- Sclaunich, C. (2020): “Coronavirus, l’Italia ha dichiarato lo stato di emergenza sanitaria per sei mesi. L’ospedale Spallanzani: «2 casi confermati, 32 sotto osservazione»” (https://www.corriere.it/cronache/20_gennaio_31/coronavirus-italia-stato-di-emergenza-news-e8c205c2-4402-11ea-b4ca-26f0f6d5d911.shtml -- last accessed: 12.11.2021).
- Sánchez, H., Álvarez, C. (2020): “Estas son las cuatro fases de la desescalada en España aprobadas por el Gobierno” (https://elpais.com/sociedad/2020-04-28/estas-son-las-cuatro-fases-para-volver-a-la-normalidad-aprobadas-por-el-gobierno.html -- last accessed: 12.11.2021).
- Severgnini C. (2020): “Coronavirus, primi due casi in Italia «Sono due cinesi in vacanza a Roma»” (https://www.corriere.it/cronache/20_gennaio_30/coronavirus-italia-corona-9d6dc436-4343-11ea-bdc8-faf1f56f19b7.shtml -- last accessed: 12.11.2021).
- Valdés, I. (2020): “Ayuso anuncia el fin de todas las limitaciones de aforo en interiores y exteriores en Madrid a partir del lunes” (https://elpais.com/espana/madrid/2021-09-29/el-lunes-4-de-octubre-se-eliminaran-las-restricciones-de-aforo-por-la-covid-en-madrid.html -- last accessed: 12.11.2021).
- Voltattorni, C. (2020): “Coronavirus, il governo: stop al ballo in tutta Italia. Stretta sulle mascherine nei luoghi della movida” (https://www.corriere.it/cronache/20_agosto_16/coronavirus-governo-discoteche-chiuse-tutta-italia-stretta-mascherine-deaa12b4-dfd4-11ea-b249-6fbea5975045.shtml -- last accessed: 12.11.2021).
- Wiebe, J. (1990): “Identifying Subjective Characters in Narrative”, in: *COLING 20*, 401 – 406.
- Wojcik, S., Hughes, A. (2019): “Sizing Up Twitter Users” (https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ -- last accessed: 15.11.2021).
- World Health Organization (2020): “COVID-19 Public Health Emergency of International Concern (PHEIC) Global research and innovation forum” (https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum -- last accessed: 12.11.2021).
- Wu, J. (2017): “Introduction to Convolutional Neural Networks”, in: *Computer Science*, n. pag.

- W3C (2021): “Extensible Markup Language (XML) 1.0 (Fifth Edition). W3C Recommendation 26 November 2008” (https://www.w3.org/TR/xml/ -- last accessed: 12.11.2021).

10 List of Figures

List of Figures

11 List of Tables

12 Declaration of Authorship

Degree programme: Language, Literature, Culture (M.A.)
Name: Jordany Werzner Regalado

I hereby declare that I am the sole author of this thesis and that I did not use any other aids or resources than the ones stated. Those parts of the paper that were taken from other works, either as quote or paraphrase, are marked by respective statements of sources. This is also the case for drawings, sketches, illustrations, and other similar works.
Furthermore, I declare that this thesis has not been handed in for a different academic assessment by me or another person. I agree to my thesis being screened using anti-plagiarism software and will also hand in a digital copy.

Giessen, 19.11.2021 Signature…………………………..

Ich erkläre hiermit, dass ich die Thesis selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, sind durch Angaben und Quellen kenntlich gemacht. Dies gilt auch für Zeichnungen, Skizzen, bildliche Darstellungen und dergleichen. Weiterhin erkläre ich, dass die Abschlussarbeit (Thesis) noch nicht im Rahmen einer staatlichen oder anderen Prüfung (z. B. als Magister-, Diplom- oder Staatsexamensarbeit) eingereicht wurde. Mit der Überprüfung meiner Abschlussarbeit mittels einer Anti-Plagiatssoftware bin ich einverstanden und reiche die Abschlussarbeit auch in digitaler Form ein.

Gießen, den 19.11.2021 Unterschrift…………………………..