# Revealing Public Opinion Sentiment Landscape: Eurovision Song Contest Sentiment Analysis

**Klara Kozolić** and **Gaurish Thakkar** and **Nives Mikelić Preradović**
Faculty of Humanities and Social Sciences,
University of Zagreb, Zagreb 10000
kkozolic@m.ffzg.hr, gthakkar@m.ffzg.hr, nmikelic@m.ffzg.hr

## Abstract

This paper discusses the use of sentiment analysis to evaluate public opinion towards the Eurovision Song Contest of 2022 and 2023 using social media comments in English, Croatian and Spanish. The study aims to determine the sentiment expressed in the comments and analyse the distribution of positive, negative, and neutral sentiments. The paper also reviews prior research on sentiment analysis related to the Eurovision Song Contest and outlines the research questions and methodology used in this study. The authors hypothesise that there are differences in how the Eurovision Song Contest is perceived by different language speakers and on different social media platforms. The methodology involves identifying relevant social media comments, studying the Eurovision timeline, collecting and preprocessing the data, and performing sentiment analysis and named entity recognition. The paper concludes by summarising the key contributions of the work and discussing future research.

## 1 Introduction

Social media comments have transformed into one of the main channels for opinion expression. This information is a great indicator of the public's opinion on numerous topics or events. Quantitative information, such as the number of comments, retweets, and likes that authors can view, supplements the textual and audiovisual content. Social media has always been essential in providing a place for discussing public thoughts and expressing views on publicly broadcast entertainment events.

The analysis of sentiment in tweets has previously been explored as a means of forecasting the outcome of certain events, such as the stock market's rise and fall (Pagolu et al., 2016). Additionally, there has been exploration of the relationship between the physical performance of basketball players and the sentiment expressed in their tweets (Xu and Yu, 2015).

The Eurovision Song Contest is a competition for the best song in European countries, held annually in May since 1956. During the event, fans post on social media content related to the competition. This provides an opportunity to evaluate methods for evaluating the sentiments of the public regarding the Eurovision Song Contest. Digital humanities strive to perform research in the humanities field with the use of technology (Svensson, 2010). In that sense, this study is related to the digital humanities, as sentiment analysis is used to extract an opinion on a social event. This paper aims to perform sentiment analysis of comments in English, Spanish, and Croatian for the Eurovision Song Contests of 2022 (ESC2022) and 2023 (ESC2023) to determine the sentiment expressed in the text. We analyse the sentiment distribution in order to measure the extent of positive, negative, and neutral subjectivity conveyed in the text. This allows for a more comprehensive assessment of the collective sentiment towards the contest and related elements.

The subsequent sections of this paper are organised in the following manner: Section 2 provides an overview of prior research conducted on the data connected to the Eurovision Song Contest. Section 3 enumerates the research questions we aim to explore. Section 3 outlines the methodology utilised in our study. The details of data preparation and preprocessing steps are described in Section 5. Section 6 of the paper provides the outcomes of the experiments, offering a thorough analysis. In Section 8, the key contributions of the work are summarised, and future research is discussed.

## 2 Related Work

There are several previous studies related to modelling various aspects associated with Eurovision. In 2019, Demergis (2019) analysed tweets in English and Spanish collected during the 2019 Eu-

rovision Song Contest to identify sentiment and rank target performers. The ranking was then compared with the ranking derived from the televoting during the Eurovision Grand Final. Kumpulainen et al. (2020) utilised sentiment analysis tools to analyse more than a million tweets in order to forecast the outcomes of ESC televoting. The study established a correlation between the sentiment ratings of tweets by comparing predicted ranks with the ranks obtained from televoting. Koski and Persson (2017), utilised the AFINN word lexicon to examine the tweets related to *Melodifestivalen*, an annual Swedish music competition organised by the Swedish national public TV. The authors employed several ranking methodologies on the participants and compared them with the actual ranks in the real world. García and Tanase (2013) utilised historical data from both Wikipedia and the official website of the Eurovision Song Contest to examine the cultural connections between European countries. The authors introduced a quantitative metric known as the Friend-or-Foe coefficient. A metric that uncovers the asymmetrical positive and negative connections among European countries indicates a correlation between cultural dissimilarity and voting biases.

## 3 Research Questions

Sentiment analysis of social media comments about Eurovision can give insight into many aspects of the song contest and its presence on social media. The research questions that we want to answer are the following:

1. What is the distribution of sentiments in the comments for each corpus? What is the distribution of sentiments in the comments for each social media platform?

2. What is the number of mentions of each participant country in each corpus?

3. Which country or artist are the negative comments most common about for each ESC year in each corpus?

4. Which country or artist are the positive comments most common about for each ESC year in each corpus?

We hypothesise that there were differences between how ESC2022 and ESC2023 appeared to different language speakers and on various social media platforms. By answering our research questions, we can get a clear idea of these differences and verify if they were in accordance with the turn of events for each ESC year.

## 4 Methodology

The method described in more detail in the paper consists of the following steps:

1. Identifying social media comments about ESC2022 and ESC2023. The first step in identifying social media comments was crucial before data collection. For Twitter, it was important to see with which hashtags we could get the most relevant data that fit our initial criteria. And for Reddit, we had to find relevant subreddits and threads about the topic of ESC2022 and ESC2023.

2. Studying the Eurovision timeline to learn when the most comments are generated. The information about the timeline of big events before, during, and after ESC helped with determining the time frame during which the comments would be collected. For example, what were the dates of the semifinals and the final, the dates of press conferences or national song selections, etc?.

3. Collecting the comments from Twitter, Reddit, and YouTube that fit the timeline and the requirements that were set.

4. Preprocessing the data. The scraped data needed to be treated differently for each social media platform. For Twitter and YouTube, we had to manually clear out the comments in other languages or comments that were not talking about the topic of ESC. Each remaining comment was manually tagged with the appropriate target language: En, Hr, or Es. For Reddit, we also had to manually clear out the comments that led to discussions that diverged too much from the topic of ESC.

5. Constructing the corpora. Corpora was constructed with Python, with the help of language tags.

6. Performing sentiment analysis on the corpora with ML models.

7. Performing Named Entity Recognition (NER) on the corpora with Python.

8. Extracting the value from and analysing the tagged data.

To answer the questions of this research, the following natural language processing and programming tasks will be performed:

- Sentiment analysis
- Named entity recognition
- Regular expressions (RegEx)

To obtain the number of mentions for each country in each corpus, we used NER and RegEx in Python. NER was performed using pretrained pipelines for each language, using publicly available spaCy models hr_core_news_sm for Croatian[1], es_core_news_sm for Spanish[2] and en_core_web_sm for English[3]. We counted the identified entities found for the country using Python. We have also performed the same task using RegEx with Python to find mentions of countries. We were able to find more mentions using RegEx than NER, but the overall results for the analysis were the same.

The data for this project consists of social media comments in English, Spanish, and Croatian for the Eurovision Song Contests of 2022 and 2023. Section 5 describes the process of data extraction and preprocessing in detail.

## 5 Data: Extraction and Preprocessing

The data for this project was collected between February and May 2023. The data that was chosen for the research was aimed at social media comments in English, Spanish, and Croatian for the Eurovision Song Contests of 2022 and 2023.

The Croatian language was chosen as our mother tongue. The English language was chosen as the most popular language on the Internet and it's the most popular foreign language in Croatia, with the largest percentage of primary school, high school and university students studying it (Kapović, 2022). Spanish language is chosen as the language whose culture is popular in Croatia thanks to music and TV series and it has a growing number of students in formal education (Kapović, 2022; Urquijo Sánchez, 2021). The data was divided into three corpora to obtain the most correct sentiment analysis results, using a different model for each language.

### 5.1 Scraping tools

For the purposes of scraping the social media comments, we have used the following three scrapers in Python:

- Twitter Scraper Selenium[4] is a Selenium-based tweet scraping tool. The tool is provided under MIT licence and allows the scraping of public tweets from a specific user profile or a hashtag.
- Universal Reddit Scrapper (URS)[5]: The tool is provided under MIT licence and allows scraping subreddits, redditors, and submission comments.
- YouTube Comment Scraper [6]: The tool is provided under Apache licence 2.0 and allows the scraping of YouTube comments under videos.

### 5.2 Data preparation and processing

The scraping for each social media site was done in batches, and each website required a different approach, which will be described in the following sections.

#### 5.2.1 Twitter

Using twitter-scraper-selenium, we have scraped the data under hashtags *"#ESC2022"* and *"#ESC2023"* to obtain the data about the target Eurovision Song Contests. We have collected the data from January until June 2022 for ESC2022 and from January to June for ESC2023. We have collected the biggest batch of tweets for the dates of semi-final one, semi-final two, and final for each of the contests. The tweets before these dates still contained a great number of opinions and news about the national selections, the Eurovision contests, the songs, and the artists, and the tweets after these dates were a great indicator of reactions to the results of the contests. After the scraping, we needed to filter out the comments in other languages and manually tag the language of the comments in order to facilitate their classification into language corpora. The comments that used the hashtags #ESC2022 and #ESC2023 but were

[1] https://spacy.io/models/hr
[2] https://spacy.io/models/es
[3] https://spacy.io/models/en
[4] https://pypi.org/project/twitter-scraper-selenium/
[5] https://github.com/JosephLai241/URS
[6] https://pypi.org/project/youtube-comment-scraper-python

not talking about Eurovision were also manually deleted.

#### 5.2.2 Reddit

Using URS, we have scraped the comments about ESC2022 and ESC2023 from the threads in the *"r/croatia"*, *"r/hrvatska"*, *"r/eurovision"* and *"r/spain"* subreddits.

Since the subreddits were in the respective languages, we did not have to classify the languages, but we still had to manually clear the comments in which users completely diverged from the topic of Eurovision.

#### 5.2.3 YouTube

Using youtube-comment-scraper-python we have scraped the comments under live broadcasts of Eurovision and of the Croatian song selection festival, Dora.

Like for Twitter, we needed to manually clear the comments in other languages and tag the language of the comments we were left with for easier classification into language corpora.

### 5.3 Creation of the language corpora

The language corpora were generated in CSV format using Python, incorporating the language tags derived from the scraped data. At the end, we had 2778 comments in Croatian, 1811 comments in Spanish, and 7613 comments in English. These corpora were used for sentiment analysis, with models for each language.

| Lang | Tokens | Types | Lang Distribution |
|---|---|---|---|
| English | 251009 | 15829 | 62.39% |
| Croatian | 46136 | 11217 | 22.77% |
| Spanish | 55861 | 8762 | 14.84% |

Table 1: The number of tokens and types and language distribution

### 5.4 Sentiment analysis

The purpose of the sentiment analysis is to determine the subjective classification of the given text, that is, to identify if a post is positive, negative, or neutral. The sentiment classification was performed using publicly available models from Hugging Face. We have employed a multilingual model for sentiment analysis: Twitter-XLM-roBERTa-base[7] for Spanish, Twitter-roBERTa-base[8] for English, and Cro-Frida[9] for Croatian. The models utilised are built upon state-of-the-art transformers (Vaswani et al., 2017; Conneau et al., 2020) architecture and have undergone fine-tuning specifically for sentiment analysis in their respective languages.

The sentiment classification was performed using publicly available models from Hugging Face. We have used a multilingual model: twitter-XLM-roBERTa-base[10] for sentiment analysis in Spanish, Twitter-roBERTa-base[11] for sentiment analysis in English and Cro-Frida[12] in Croatian. The models utilised are built upon state-of-the-art transformers (Vaswani et al., 2017; Conneau et al., 2020) architecture and have undergone fine-tuning specifically for sentiment analysis in their respective languages.

The examples of sentiment classified data (negative in these cases) for each corpus are:

- *Unpopular opinion: Sweden performance doesn't deserve to win. Too many mistakes. #Eurovision #Sweden #EUROVISION #ESC2022*, negative (0.8983), neutral (0.0916), positive (0.0101)
- *E ovo je bilo dosadno. Ne ponovilo se #ESC2023*, 0
- *Loreen ha ganado pero sintiéndolo mucho no se lo merecía y no pienso discutir esto #ESC2023 #EUROVISION #EUROVISION2023*, [{'score': 0.5728, 'label': 'negative'}]

## 6 Results and Analysis

To evaluate the corpora, we have randomly selected 300 comments from each corpora and tagged their sentiment manually. Tagged corpora were compared to corpora containing sentiment predictions using Python. The Table 2 shows the precision, recall, F-1 measure and accuracy of the models used for sentiment analysis.

[7] https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment

[8] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

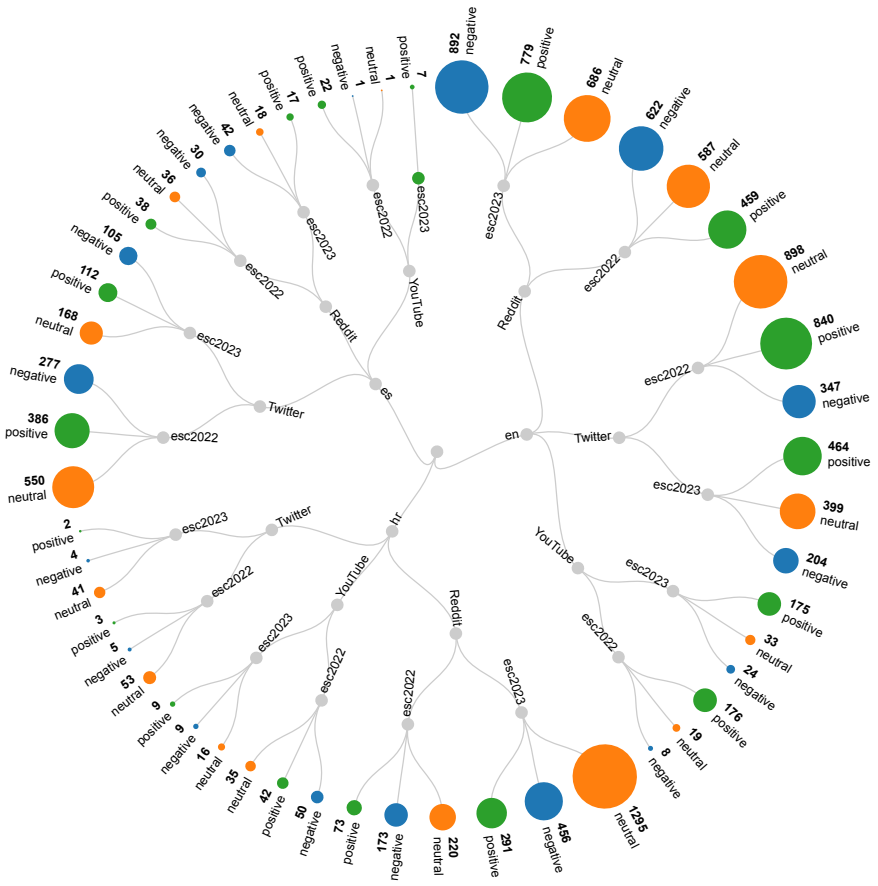
[9] https://huggingface.co/thak123/Cro-Frida

[10] https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment

[11] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment

[12] https://huggingface.co/thak123/Cro-Frida

Figure 1: Corpora categorised based on their language, social media source, year, and sentiment label.




| Lang | P | R | F-1 | Acc |
|---|---|---|---|---|
| English | 67.21 | 66.78 | 66.53 | 66.78 |
| Croatian | 65.27 | 62.58 | 62.04 | 62.58 |
| Spanish | 0.7453 | 74.67 | 74.50 | 74.67 |

Table 2: Evaluation of sentiment analysis approaches employed for categorising tweets. P:precision, R:recall, Acc: accuracy.

| Lang | Year | Pos | Neg | Neu |
|---|---|---|---|---|
| English | 2022 | 1475 | 977 | 1504 |
| | 2023 | 1418 | 1120 | 1118 |
| Croatian | 2022 | 118 | 228 | 308 |
| | 2023 | 302 | 469 | 1352 |
| Spanish | 2022 | 446 | 308 | 587 |
| | 2023 | 136 | 147 | 186 |

Table 3: The distribution of sentiments in each corpus

### 6.1 The distribution of sentiments in the comments for each corpus

After the task of sentiment analysis, the following sentiment distribution was found:

Table 3 shows sentiment distribution for each Eurovision year in each corpus. For the English corpus, we can note that there were the most neutral comments for ESC2022 and the most positive comments for ESC2023. Taking into account that there were a lot more positive than negative comments for ESC2022, we can conclude that both of these contests were mostly perceived as positive. For Croatian, each Eurovision had the greatest number of neutral comments. As opposed to English, we can note a higher number of negative than positive comments. Thus, we can conclude that both ESC2022 and ESC2023 were perceived as negative. For Spanish, each Eurovision had the greatest number of neutral comments as well. ESC2022 had more positive than negative comments, and ESC2023 had more negative than positive com-

ments. This coincides with the good placement of Spain in 2022 and the bad one in 2023.

### 6.2 The distribution of sentiments in the comments for each social media

The distribution of sentiments in the comments for each social media can be seen in Table 4.

| Social Media | Year | Pos | Neg | Neu |
|---|---|---|---|---|
| Reddit | 2022 | 570 | 825 | 843 |
| | 2023 | 1087 | 1390 | 1999 |
| Twitter | 2022 | 1229 | 629 | 1501 |
| | 2023 | 578 | 313 | 608 |
| YouTube | 2022 | 240 | 59 | 55 |
| | 2023 | 191 | 33 | 49 |

Table 4: The distribution of sentiments in each social media

On Reddit, each Eurovision had the most neutral comments. Each Eurovision also had more negative than positive comments, which leads to the conclusion that both years were perceived as negative. Twitter shows the opposite situation, with the most neutral but more positive comments. Thus, we can conclude that both years were perceived as more positive. On YouTube, the greatest number of comments were positive for each Eurovision. These results indicate that there was more critical discussion regarding various aspects of ESC2022 and ESC2023 on Reddit than on Twitter and YouTube. This fits the format of Reddit as one of the most popular sites for opinion exchange and discussion.

### 6.3 The number of mentions of each participant country in each corpus

Figure 2 shows the number of mentions of each country participating in the Eurovision in 2022 and 2023 for the English corpus. For ESC2022, the most mentioned country is Ukraine, the winner of that year's edition. We can see Spain and the UK mentioned the most after Ukraine. The UK took second place and offered to host in place of Ukraine in 2023. Spain did well with the juries and televotes, placing third. The least mentioned countries are the ones that did not qualify for the final or were placed in lower positions, typically out of the top 10. For ESC2023, the most mentioned country is Finland, a runner-up for 2023. Finland was a fan favourite and received the most votes from the audience. The second country is Sweden, the winner of ESC2023. The least mentioned countries are the ones that did not qualify for the final or were placed in lower positions, typically out of the top 10, except for Germany and Spain, which ended up with some of the worst results, which did not sit well with the audience. Additionally, three countries dropped out of the competition in 2023 and were not mentioned at all for that year: Montenegro, North Macedonia, and Bulgaria.

Figure 3 shows the number of mentions of each country participating in the Eurovision in 2022 and 2023 for the Spanish corpus. The Spanish corpus did not reflect the general opinion about Eurovision as much as the opinions of Spanish people. In both Eurovision song contests, Spain was the most mentioned country. Spain was mentioned more in ESC2022 when its entry placed third, and many fans thought it would have won if it hadn't been for the war in Ukraine. The higher number of mentions of other countries for ESC2022 also indicates that ESC2022 was more viewed than ESC2023 in Spain.

Figure 4 shows the number of mentions of each country participating in the Eurovision in 2022 and 2023 for the Croatian corpus. Like in the Spanish corpus, the Croatian corpus reflects the opinions of Croatians regarding ESC2022 and ESC2023. The most mentioned countries, alongside Croatia, were the countries that placed higher in each ESC edition. As opposed to Spain, the bigger interest for Croatian data is in ESC2023, where Croatia qualified for the finals for the first time since 2017.

### 6.4 Country: negative comments

The extraction of mentions of each country and sentiments related to that country showed insight into which country faced the most negative comments in each corpus. Starting with the English corpus, the country with the most negative comments associated with it for ESC2022 was Ukraine. This result was not surprising, as Ukraine won in 2022. Many people believe it was a political win due to the Russian invasion of Ukraine that started in 2022 and because of which Russia was banned from participating in ESC2022 (Welslau and Selck, 2023). For ESC2023, the country with the most negative comments was Finland, the runner-up for that year. Following ESC2023, there was a lot of discussion involving Sweden and Finland, where fans were arguing that Finland was the audience favourite and should have won, saying that the juries wanted Sweden to win. These discussions led both Finland and Sweden to be mentioned in both a negative and positive light, the most. The countries with the

Figure 2: Frequency distribution of nations cited in the English corpus.

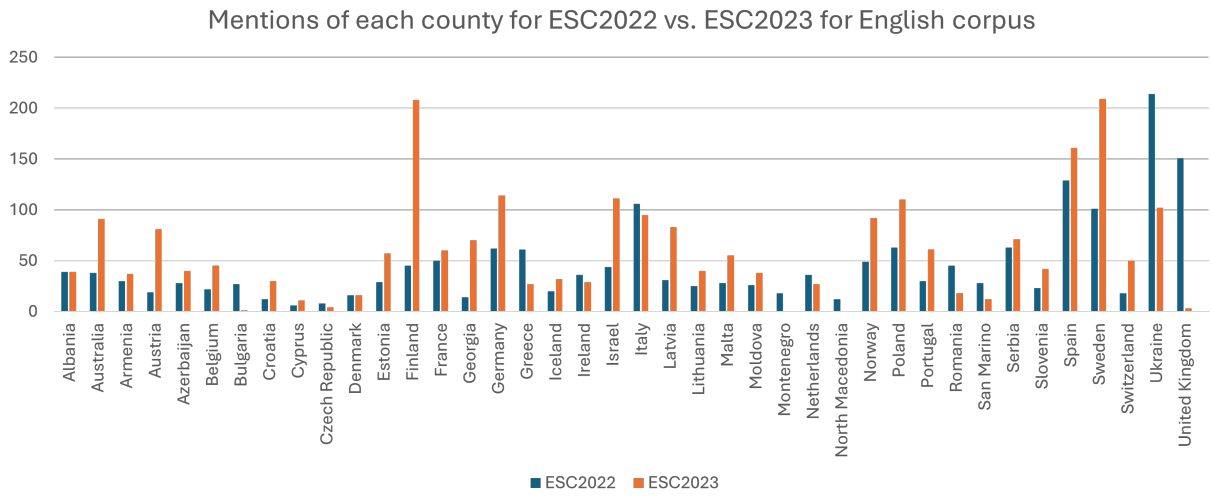
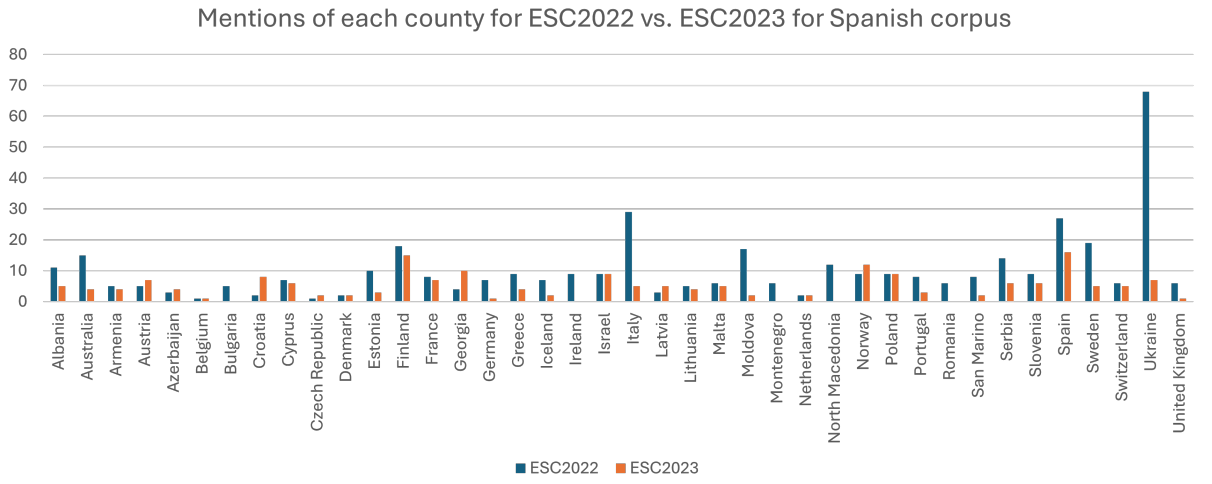



Figure 3: Frequency distribution of nations cited in the Spanish corpus

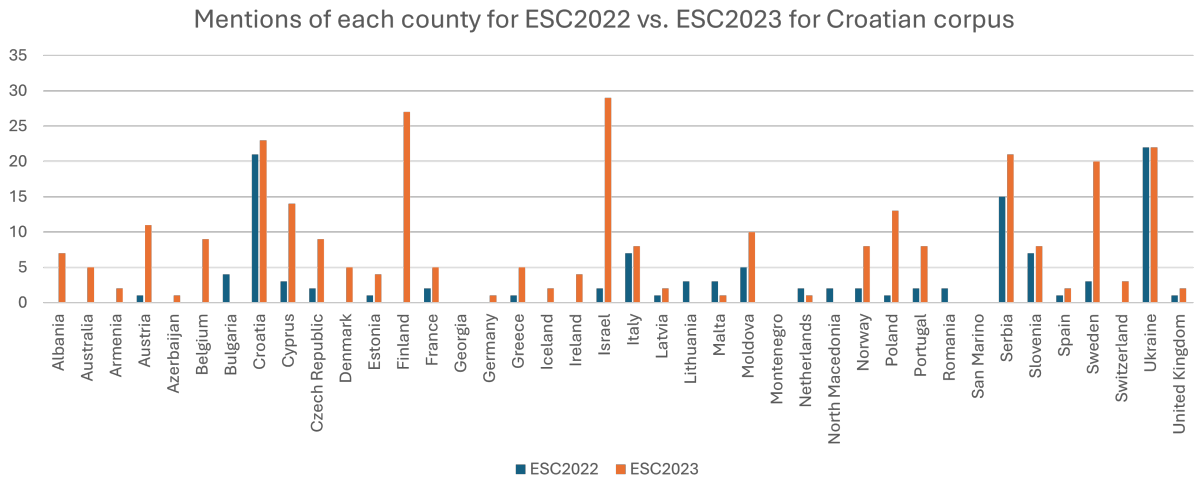



most negative mentions in the Croatian corpus for ESC2022 were Croatia and Ukraine. Croatia sent a song that didn't qualify, and the public in Croatia did not like it, while the number for Ukraine could be explained by its triumph, which many perceive as political. For ESC2023, the most negatively mentioned countries in Croatian corpus were Finland and Serbia. Finland could be explained with the same reasoning as for the English corpus: the Finland vs. Sweden discussions. As for Serbia, neighbouring countries or countries that have something in common, such as Slavic roots in this case, tend to mention one another more in both a positive and negative light. In the Spanish corpus, the most negatively mentioned country for ESC2022 was Spain. These comments can be explained by the fact that Spain placed third, and the Spaniards believed it should have won, so they were expressing their dissatisfaction. For ESC2023, the most negatively mentioned country was Spain again. Spain received very few points from the public in 2023 and the public in Spain and worldwide didn't seem to receive the modern rendition of flamenco well.

### 6.5 Country: positive comments

The extraction of mentions of each country and sentiments related to that country showed insight into which country faced the most positive comments in each corpus. In the English corpus for ESC2022, the country with the most positive comments was Ukraine. While it generated the most negative comments, it also generated the most pos-

Figure 4: Frequency distribution of nations cited in the Croatian corpus




itive ones, since a lot of people expressed their solidarity with Ukraine and liked the song they performed. For ESC2023 in the English corpus, the country with the most positive mentions was Finland. However, it's important to note that Sweden had only 4 comments marked as positive, fewer than Finland, yet it achieved a higher positive average score. This demonstrates really well how people had polarising opinions and discussions on which of the two countries should have won the ESC2023. In the Croatian corpus for ESC2022, the most positively mentioned countries were Croatia and Poland. ESC2022 was not as commented on and followed as ESC2023 in Croatia, ending with Croatia being mentioned as the most positive and most negative. For ESC2023, the most positively mentioned countries were Croatia, Finland, Israel, and Ukraine. All four countries were discussed on social media and had both positive and negative comments. In the Spanish corpus for ESC2022, the most positively mentioned country was Spain, which placed third that year, and the comments were convinced of Spain's win. For ESC2023, the most positively mentioned countries were Israel and Georgia. Israel placed third in 2023 and was a very discussed entry for that year, while Georgia didn't qualify for the finals and many people believed it should have.

## 7 Discussion

After the analysis of the data, we saw that each corpus had a great number of comments tagged as neutral. If we take into account only positive and negative comments, the distribution is the following: for the English corpus, the sentiment was positive for both ESC years; for the Croatian, the sentiment was negative for both ESC years; and for the Spanish, the sentiment was positive for ESC2022, and it was negative for ESC2023. The distribution of sentiments for each social media platform revealed that Twitter and YouTube had more positive comments, and that Reddit had more negative comments for each ESC year. The number of mentions for each participant country followed the outcome of each ESC year in the English corpus, with the most mentioned countries being the countries that scored higher or qualified for the final. The mentions of countries in Croatian and Spanish corpora also followed the outcome of ESC, but with more mentions of their own countries or their neighbouring countries. The negative comments about countries revealed that in each corpus, the most negatively mentioned countries were the ones whose victory was controversial or those who people felt deserved to win. In Croatian and Spanish corpora, we again saw the focus shifting more to their own or neighbouring countries. There is a similar situation with positive comments about countries, where the most discussed countries had more positive mentions than the others, with more details about their own countries in Croatian and Spanish comments. As demonstrated in the analysis, Eurovision can reflect the cultural and political situation in Europe and its entries, and the public can use it as a means of social commentary (López Zapico, 2023). Its cultural importance is manifested through the sense of unity of Europe and belonging

that it's aiming to emit to its viewers through the practices of so-called "postcards", videos before each performance that show the artist and natural beauty of each country (Coupe and Chaban, 2019). In their work (Spierdijk and Vellekoop, 2006) have found that not only geographical proximity plays a role in Eurovision votes, but also shared cultural background and cultural similarities such as a language result in genuine votes. In our data, we have found that countries with geographical proximity, language, shared culture and history tend to mention one another more in both positive and negative contexts. If we take Spain as an example, we can see that it favours Italy, but dislikes France. What it has in common with these two counties is the same language group and cultural and historical ties. The same can be said for Croatia that either favours or dislikes Serbia, its neighbouring country with shared cultural patterns and language group. The same was found in a study by (Spierdijk and Vellekoop, 2006) where, for example, Croatia was favoured by Slovenia, but disliked by Hungary, both neighbouring countries or Spain that favoured Italy.

## 8 Conclusion

This research gave an insight on how sentiment analysis can demonstrate public opinion towards the Eurovision Song Contest. The analysis was performed on data in three different languages: English, Spanish, and Croatian. The research found that the English corpus showed general opinion very well, while the Spanish and Croatian corpora showed opinions and sentiments more concentrated on their own entries, very popular entries, or entries that have something in common with these countries, such as language group or neighbouring geographical location. The NLP tasks of sentiment analysis and NER proved to be beneficial in painting the picture of opinion towards the Eurovision Song Contest as a whole: before, during, and after the show. The results of the analysis were in accordance with how certain entries and countries were perceived during the show and the reaction to them after the show. However, a clear limitation of our analysis are the differences between languages and language resources for the three languages we chose, cultural references and sarcasm in comments, and the bias in the data, which was shown well in Croatian and Spanish examples. Furthermore, during data collection, we noticed that people from different countries used different social media sites to talk about Eurovision. For example, we saw that the Croatians were more active on Reddit, whereas the Spaniards were more active on Twitter. All of these limitations impact the end goal of sentiment analysis for this research: determining the sentiment about ESC in different countries. Another limitation that needs to be considered is the subjectivity of choosing both the languages and certain social media sites as sources. In order to increase the objectivity of the results, data from more social media or internet sites could be added. The mention of the artists or songs, the addition of more categories of sentiment for an even more detailed demonstration of the opinion, or the addition of comments in more languages are some of the numerous elements that could be added to expand this research. In conclusion, sentiment analysis of ESC-related comments can be beneficial for studying the social and political situation in Europe and how it reflects on ESC or for studying the entertainment value ESC brings to its fans each year.

## References


Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzm'an, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Advances in Neural Information Processing Systems*.

Tom Coupe and Natalia Chaban. 2019. Creating europe through culture? the impact of the european song contest on european identity. *Empirica*, 47(4):885–908.

Dimitri Demergis. 2019. Predicting eurovision song contest results by interpreting the tweets of eurovision fans. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pages 521–528.

David García and Dorian Tanase. 2013. Measuring cultural dynamics through the eurovision song contest. *Advances in Complex Systems*, 16(08):1350037.

Marko Kapović. 2022. Strani jezici u formalnom obrazovanju u hrvatskoj. *Strani Jezici*, 51(2):283–309.

Alexander Koski and Jennifer Persson. 2017. And the winner is... : Predicting the outcome of melodifestivalen by analyzing the sentiment value of tweets.

Iiro Kumpulainen, Eemil Praks, Tenho Korhonen, Anqi Ni, Ville Rissanen, and Jouko Vankka. 2020. *Predicting Eurovision Song Contest Results Using Sentiment Analysis*, pages 87–108.

Misael Arturo López Zapico. 2023. Europe's living a celebration? el impacto social y cultural de la europeización de españa a través de canciones y otros productos culturales. *Revista de Estudios Europeos*, 82:152–183.

Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. 2016. Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pages 1345–1350. IEEE.

Laura Spierdijk and Michel Vellekoop. 2006. Creating europe through culture? the impact of the european song contest on european identity. *Empirica*, 47(4):885–908.

Patrik Svensson. 2010. The landscape of digital humanities. *Digital Humanities Quarterly*, 4.

José Ignacio Urquijo Sánchez. 2021. *EL ESPAÑOL EN CROACIA*, pages 385–395.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5099–5109.

Lea M Welslau and Torsten J Selck. 2023. Geopolitics in the esc: Comparing russia's and ukraine's use of cultural diplomacy in the eurovision song contest. *New Perspectives*, page 2336825X231222000.

Chenyan Xu and Yang Yu. 2015. Measuring nba players' mood by mining athlete-generated content. In *2015 48th Hawaii International Conference on System Sciences*, pages 1706–1713. IEEE.