

Understanding Gender Differences in Online Comments about Femicides in Italy

Claudia Schiavetti

March 25, 2025

This study explores gender differences in public reactions to femicides in Italy by analyzing YouTube comments from major Italian news channels (RAI, La7, and Nove) using data collected via the YouTube API. The research investigates the emotional tone and discourse structures of male and female users in response to high-profile femicide cases. Gender is inferred through a hybrid approach using name-matching and large language models (LLMs). The study also considers activity levels (most vs. less active users) and the role of emoji in sentiment classification. Results indicate distinct gender-based patterns, with female users showing more empathetic and relational language and male users focusing more on justice-oriented and negative expressions. Findings highlight the importance of lexicon choice and textual context (e.g., negation and emoji use) in shaping sentiment analysis outcomes.

Introduction and Motivation

Femicide, the killing of women and girls because of their gender, is a serious global issue and the most extreme form of violence against women. It is rooted in misogyny, harmful beliefs, and social norms that fuel gender inequality. The scale of this problem has drawn growing attention from both the public and researchers, especially regarding how it plays out in online spaces. Social media platforms, like YouTube, offer a valuable place to observe public reactions to important social issues such as femicide.

In Italy, public debate around femicide has intensified, especially after several high-profile cases. Among these, the murders of Giulia Cecchettin and Giulia Tramontano sparked national attention and outrage. Giulia Cecchettin, a 22-year-old university student, was murdered by her ex-boyfriend Filippo Turetta in November 2023. Her disappearance and the discovery of her body, with more than twenty stab wounds, triggered nationwide protests and renewed focus on domestic violence. The severity of the crime, coupled with the young age of those involved, further fueled public anger. The case ended with Turetta receiving a life sentence, turning the event into a symbol of the broader problem of femicide in Italy.

Similarly, Giulia Tramontano, a 29-year-old woman who was seven months pregnant, was killed by her boyfriend Alessandro Impagnatiello in May 2023 near Milan. She suffered 37 stab wounds, two of which were fatal. The fact that she was pregnant added emotional weight to the case and shaped public sentiment. Impagnatiello confessed, and the case remains under public scrutiny, fueling further discussion on violence against women.

This study examines public sentiment on these cases by analyzing comments posted under related videos on the YouTube channels of RAI, La7, and Nove. RAI is Italy's national public broadcaster, offering a wide range of news and general content. La7 is a commercial broadcaster known for news and political talk shows, attracting a politically engaged audience. Nove, part of Warner Bros. Discovery, is a general entertainment channel with a diverse viewership. These three platforms provide varied perspectives on how different audiences engage with femicide cases online.

The dataset consists of YouTube comments on videos about the femicides of Giulia Cecchettin and Giulia Tramontano from these media outlets. These comments are used to explore public sentiment, with a focus on identifying differences based on the gender of the commenters. Gender is inferred from usernames using a large language model (LLM) and name-matching techniques. The goal is to examine whether female and male commenters express sentiment differently when reacting to news about femicide.

This leads to the key research question: Is there a significant difference in sentiment between female and male commenters on YouTube videos about the Giulia Cecchettin and Giulia Tramontano femicide cases published by RAI, La7, and Nove?

This question is important for several reasons. While previous research has explored gender differences in online communication and sentiment analysis, little work has focused specifically on how gender shapes reactions to femicide in Italy on social media. Studies suggest women might express emotions more strongly and use more civil language online, while men might engage more frequently in less filtered or negative commentary. However, how these general patterns apply to sensitive topics like femicide requires dedicated investigation.

Additionally, this research makes a methodological contribution by using a large dataset of user-generated comments to study public reactions to social issues. The comment sections of news videos give a direct and often raw look at public opinion, making them valuable for social research. By combining Natural Language Processing (NLP), sentiment analysis, and gender classification using an LLM, this study offers a new approach to understanding gendered patterns in online discussions.

Finally, this work aims to deepen understanding of how men and women engage with sensitive issues like femicide in Italy. Identifying differences in sentiment can provide insights into varying levels of empathy, awareness, or perspectives on violence against women. The findings could help shape future research on gender and media, as well as improve public awareness campaigns aimed at addressing femicide.

This study is motivated by the persistent issue of femicide in Italy, where legal measures have not stopped the alarming number of women killed by partners or ex-partners. Data from the Italian Ministry of the Interior show that one femicide happens almost every three days, raising questions about why this violence continues and how society responds.

Public protests and feminist movements have been key in pushing the issue of “femminicidio” into national discussions, framing it as a problem rooted in patriarchal systems and gender inequality. Media coverage and social mobilization following the cases of Cecchettin and Tramontano reflect how much this issue resonates with the public.

In today’s world, platforms like YouTube are central to public debate. Comments on news videos show how people react to major events, often without filters. Analyzing these comments helps reveal public sentiment and offers insights into how people process emotionally charged topics like femicide.

This study focuses on possible differences in how men and women react online to violence against women. Past research shows that gender may influence online communication styles and emotional expression. For example, women may show more emotional responses and civility in online discussions, while men may display more aggressive or dismissive behaviors. Understanding these patterns in the context of femicide could shed light on deeper social attitudes and biases.

This research contributes to several areas: Social media analysis, using YouTube comments to study public sentiment around violence against women. Gender studies, exploring gendered patterns in online reactions to femicide. Media studies, comparing responses across different types of media outlets.

Ultimately, this project aims to provide a clearer picture of how gender shapes online reactions to femicide in Italy, contributing to a broader understanding of this urgent social issue.

Methodology

Data Collection

The target population of this study comprises YouTube commenters who react to and engage in discussions about femicides in response to videos covering these cases, posted by the official channels of major Italian television networks. The dataset includes user-generated comments on YouTube videos related to two highly famous femicide cases in Italy in 2023: the murders of Giulia Cecchettin and Giulia Tramontano. Comments were collected from the official YouTube channels of three leading Italian broadcasters: RAI, La7 (including sub-channels such as TgLa7 and La7 Intrattenimento), and Nove.

Data collection was performed using the YouTube API through the *tuber* R package. The selected videos include all content published on the official channels within six months following the victims' deaths. The search strategy was based on the victims' full names as keywords, followed by filtering the results according to the defined time windows. Specifically, for Giulia Cecchettin, the observation period spans from November 11, 2023, to May 11, 2024, while for Giulia Tramontano, the window ranges from May 27, 2023, to November 27, 2023.

To ensure dataset consistency and focus, only first-level comments, those posted directly in response to the videos, are retained, while replies to other comments are excluded. This filtering is based on the *parentID* variable, where first-level comments are marked by an NA value, while second-level comments display the ID of the parent comment. Additionally, only videos with at least one comment are included.

It is important to note that not all channels published videos that meet both the keyword and time frame criteria. However, this does not affect the analysis, as the two cases are treated collectively under the broader category of femicides, rather than as isolated incidents.

The comments were first saved into different datasets and then were merged into a final dataset, resulting in a comprehensive collection of public reactions to the femicides.

The final dataset contains 40224 observations and 17 variables. The most relevant variables from the resulting dataset are:

- **ChannelID** and **source_file**: Both variables capture the origin of the data by identifying the official YouTube channel from which each comment is retrieved. While *ChannelID* records the native alphanumeric identifier assigned by the YouTube platform, *source_file* is generated to facilitate interpretability by mapping each *ChannelID* to a recognizable label corresponding to the broadcaster's channel. Specifically, *source_file* is a factor variable categorizing each observation by the respective channel name. This variable is created during the data integration process by appending an identifier to each set of comments when the datasets, originally collected separately by channel, are combined into a single unified dataset. This step facilitates subsequent data handling and improves clarity during the analytical process.
- **textOriginal**: This is a text variable that stores the full content of each user comment.
- **authorDisplayName**: This is a text variable containing the username of each commenter. It is later used in the gender assignment procedure described in the following section.

- **likeCount:** This is a numeric variable that records the number of likes received by each comment.
- **parentId:** As previously mentioned, this variable has two possible outcomes: an NA value, indicating that the comment is a first-level comment (posted directly under the video), or the ID of the parent comment if it is a reply (second-level comment). This variable is used to filter and retain only first-level comments for the analysis.

Gender Assignment

A key methodological challenge in this study is the gender assignment of YouTube commenters, as the YouTube API does not provide gender information due to privacy restrictions. However, gender attribution is essential for the scope of this research. To address this, a gender assignment procedure is implemented using a combination of traditional name-matching techniques and a Large Language Model (LLM). The procedure is structured into three main steps:

Firstly, a dataset of common Italian first names and their associated genders is created. This list functions as a reference for identifying potential names embedded within usernames. Usernames are systematically scanned to extract any substring that corresponds to a name listed in the reference dataset. For example, in the username “Marco_1960”, the string “Marco” would be extracted and linked to the male gender based on the name list.

Secondly, when a match is identified between a username and the reference dataset, the corresponding gender is directly assigned to the commenter. This step is crucial because it reduces the computational burden of using LLMs by focusing initial efforts on easily identifiable usernames. In doing so, the LLM is applied only to the more complex or ambiguous usernames that cannot be resolved through name-matching alone.

Lastly, for the remaining unmatched usernames, an LLM is used to infer the most likely gender. This is performed using the OpenAI API through the *ellmer* R package. To further optimize computational efficiency, the LLM is applied exclusively to the set of unique usernames. In cases where a commenter appears multiple times in the dataset (i.e., posts multiple comments), gender prediction is conducted once for the unique username and subsequently assigned to all corresponding observations.

To partially address the limitations of this method, an additional “unknown” category is introduced. When no confident gender match can be established—either through the name-matching step or the LLM—the gender is labeled as “unknown.” This helps mitigate risks related to forced or uncertain classifications.

The code is first run on a subset of 100 obs of the dataset and an evaluation of this method is performed to verify the efficiency of the assigning gender procedure.

Table 1: Confusion Matrix

Metric	male	female	unknown
True Positives (TP)	29	39	18
False Negatives (FN)	0	0	5
False Positives (FP)	5	5	0
Accuracy	0.86		
95% CI	(0.7763, 0.9213)		
Sensitivity	1.0000	0.8732	0.7632
Specificity	0.9070	0.9123	1.0000
Precision (PPV)	0.7632	0.8864	1.0000

The model achieves a strong overall accuracy of 86%, indicating that most cases are classified correctly. Sensitivity is perfect for the “male” class and high for “female,” while it is lower for “unknown,” showing that the model struggles to capture some “unknown” cases. Specificity is excellent across all classes, particularly for “unknown” (1.0), reflecting a minimal false positive rate.

Precision is generally good, though slightly lower for the “male” class (0.76), suggesting that some usernames classified as “male” may belong to other categories. The model appears more inclined to assign a gender (male/female) rather than labeling users as “unknown,” likely influenced by a prompt encouraging gender predictions based on probable name associations.

In summary, the model is statistically solid and reliable for gender classification, but could benefit from improved detection of the “unknown” class to reduce over-assignment to binary gender categories. For further details, including additional performance metrics and visualizations, see the Appendix.

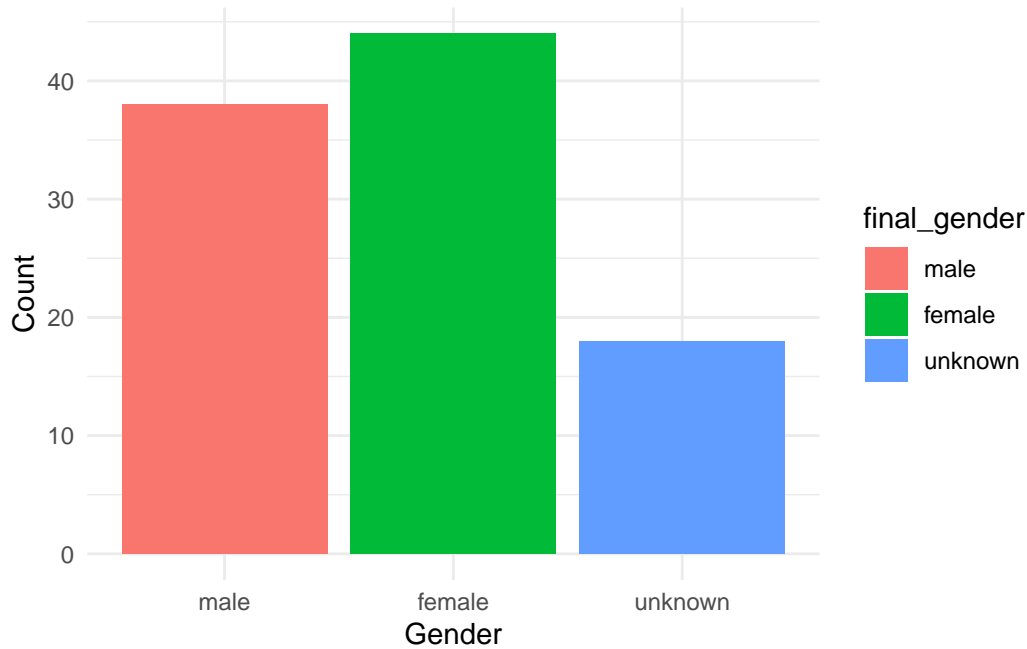


Figure 1: Distribution of Predicted Gender

The bar chart shows the distribution of predicted genders in the test dataset following the gender assignment procedure. The majority of usernames were classified as “female” (over 40 cases), followed by “male”. A smaller but still relevant portion were assigned to the “unknown” category.

This suggests that while the model tends to successfully assign a binary gender (male or female) in most cases, there is a notable fraction of usernames for which the model could not confidently infer gender, resulting in an “unknown” label. This aligns with the earlier observation that the model is somewhat conservative when uncertain, but still shows a tendency to prefer gendered classifications over “unknown.”

Since the evaluation confirmed the reliability of the method, the procedure is applied to the full dataset. To speed up the process and reduce the risk of connection errors, the dataset is split into subsets of approximately 5,000 observations. Within each subset, the procedure is run in smaller batches of 50 observations, which are then combined. The corresponding code is provided in the Appendix, and the prompt used for this procedure is also available in the Quarto file. Finally, all subsets are merged into a single dataset, which now includes a new column called **final_gender**, a categorical variable with three possible values: “female,” “male,” and “unknown.”

The final unit of analysis for this study is the individual YouTube comment. Each comment represents a unique instance of user-generated content and serves as the primary level at which data is collected and analyzed. By focusing on single comments, the study captures expressions of public sentiment, allowing for the examination of how users engage with and react to media coverage of femicide cases.

Emoji Translation

An additional challenge in the data cleaning process involves the treatment of emojis. Given that one of the objectives of this study is to conduct a sentiment analysis, it is inappropriate to disregard emojis entirely, as they convey emotional and contextual details. However, for other descriptive analyses, such as word cloud visualizations, retaining emojis could introduce noise or distort the interpretability of textual patterns.

To address this, two versions of the comment text are created. The first version excludes emojis entirely, retaining only the plain text. The second version replaces each emoji with its corresponding textual description. A key obstacle in this process is the need to generate accurate emoji descriptions in Italian, as the initial description dataset contains emoji labels in English.

The adopted procedure consists of loading the English-language emoji description dataset, which contains 5,042 observations, and subsequently translating these descriptions into Italian. The translation is performed using the OpenAI API, where the English descriptions are processed in batches of 50 observations to ensure accurate and context-sensitive translations. Full details of the procedure, including the exact prompt and workflow, are documented in the Appendix.

By checking the length of the resulting translation vector, it is observed that six translations are missing. These missing entries are manually detected and completed to ensure consistency with the formatting and structure of the existing translations.

To minimize potential bias in subsequent analyses, certain terms recurrent in the emoji descriptions, such as “woman,” “man,” “skin,” “face,” and “flag”, are removed from the text, as their inclusion could disproportionately influence sentiment analysis outcomes.

The final dataset contains two additional columns: **Erase_Emoji** and **Rep_Emoji**. The **Erase_Emoji** column contains comments where emojis have been removed, leaving only the pure text. In this version, each emoji is replaced with an empty space, which also avoids generating NA values in cases where comments consist solely of emojis. The **Rep_Emoji** column contains comments where each emoji has been substituted with its corresponding Italian description, preserving the semantic contribution of emojis for sentiment analysis purposes.

Most Active Users

During the data cleaning process, it was also observed that several comments were written by the same users. Specifically, there are 21,860 unique users out of a total of 40,224 comments. Therefore, it is useful to differentiate between highly active users and less active ones in some parts of the analysis. To this end, a new column called **active_users** was created, assigning a value of 1 to users who wrote more than 3 comments, and 0 otherwise.

Sentiment Analysis

When interpreting text, readers rely heavily on the emotional tone conveyed by words to determine whether the content is positive, negative, or expresses a more nuanced emotion such as surprise, anger, or sadness. In computational analysis, this process can be automated through text mining techniques that assess emotional content. A common approach involves breaking the text into individual tokens (words) and aggregating the sentiment scores of each token to infer the overall sentiment of a sentence or document. While several methods exist for sentiment analysis, this token-based approach is widely used due to its compatibility with many available tools.

To measure the sentiment of individual words, various sentiment lexicons (also known as dictionaries) are available. These lexicons categorize words according to emotional labels, ranging from basic positive/negative polarity to more detailed emotion categories. In this analysis, two lexicons were employed:

- *NRC Sentiment Lexicon* :Developed by the National Research Council of Canada, the NRC lexicon categorizes words based on the emotions they evoke, including positive, negative, anger, fear, joy, sadness, surprise, and trust. Each word is assigned a binary score—1 if it expresses a specific emotion, 0 otherwise. The Italian version of this lexicon, available via the *syuzhet* R package, was used for this project. The most frequently applied functions from this package were *get_nrc_sentiment()*, which returns an emotion score table for each word or sentence, and *get_sentiment()*, which provides an overall sentiment score for full sentences.
- *Italian Sentiment Lexicon*: Created by the Consiglio Nazionale delle Ricerche (CNR), this lexicon is available on the official CNR website and was imported into R using the *.lmf* format. Unlike the NRC lexicon, it assigns words into three categories: positive, neutral, and negative. Despite its simpler classification system, it has been found to perform more accurately in certain contexts than the NRC lexicon.

The code used to import and apply both lexicons is available in the appendix.

Another challenge encountered during the analysis was how to handle plurals, gender variations (masculine and feminine), and different conjugations of the same verb. To address this issue, *stemming* was applied, a common text preprocessing technique in sentiment analysis that reduces words to their root or base form by removing suffixes or prefixes. For instance, “running,” “runs,” and “ran” are all reduced to the stem “run.” This process helps consolidate word variations so they are treated as a single term, improving model performance by reducing vocabulary size and noise. To avoid displaying truncated or hard-to-read stems in the final output, the word with the highest frequency from the group is usually shown, while its frequency count reflects the combined total of all stemmed word forms.

Exploratory Analysis

In this word cloud, only words with a frequency greater than 800 occurrences are displayed. The most frequent word is “Giulia”, appearing 4,878 times, which is expected since both victims shared this name and it was widely mentioned in the comments. Other highly frequent words include “donne” (4,689 occurrences), meaning “women,” and “ragazza” (3,625 occurrences), meaning “girl,” reflecting the fact that both victims were young women. The word “figlio” (4,147 occurrences), meaning “son,” also appears frequently. This could be attributed to two reasons: many comments refer to the pain a parent feels when losing a child, supported by the high frequency of the word “dolore” (meaning “pain”), and because Giulia Tramontano was seven months pregnant when she was killed. Other notable words include “povera” (meaning “poor” but in this context has the sense of unfortunate girl), “vita” (meaning “life”), and “mostro” (meaning

7



(a) WordCloud of the female users



7



(a) WordCloud of the Less Active Users



11

cloud of the most active users displays fewer unique words, suggesting a tendency to repeat the same terms more frequently, while less active users appear to use a broader range of vocabulary. The word “Giulia” emerges as the most frequent term among the most active users, whereas “donne” (women) is the most common word among less active users. The word “uomo” (man) is notably present in the word cloud of less active users but does not appear among the most active ones. In addition, the word cloud of less active users includes more dramatic terms such as “mostro” (monster), “ergastolo” (life imprisonment), “morte” (death), and “male” (evil), which are absent from the word cloud of the most active users.

Positive and Negative Words

This section presents an analysis of the most positive and negative words within the entire dataset, as well as across each of the subgroups under consideration. The analysis is performed using both lexicons, and the text version containing the translated emoji descriptions is employed, as emojis are frequently linked to emotions and sentiment, making their inclusion in the analysis appropriate. For this purpose, a separate corpus is created and stemmed for each subgroup.

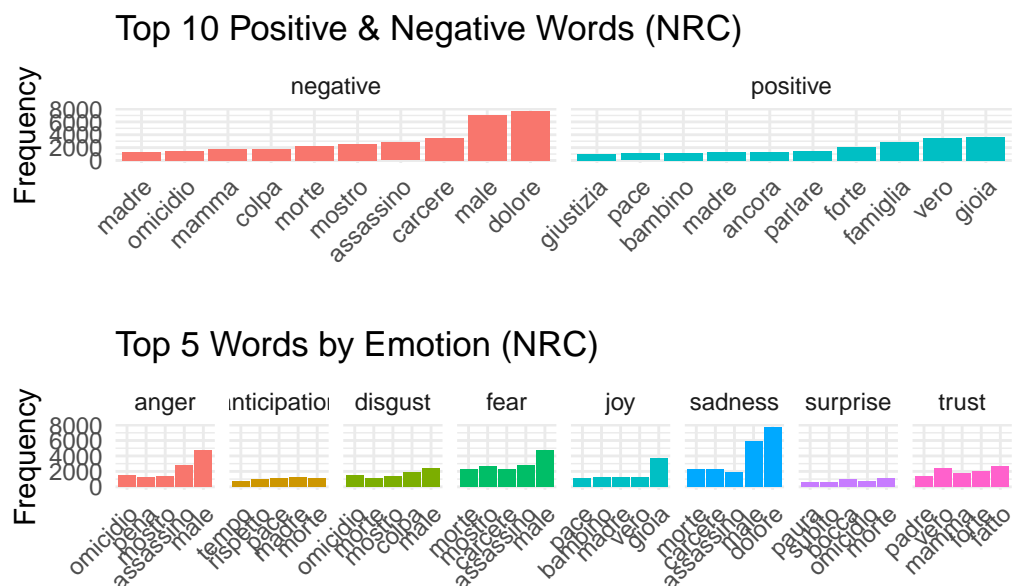


Figure 7: Word Frequencies based on the NRC Sentiment Lexicon

The first plot displays the Top 10 positive and negative words in the complete dataset. The negative words are notably more frequent, with “dolore” (pain) and “male” (evil) being the most prominent, followed by “carcere” (prison), “assassino” (murderer), and “morte” (death). This reflects the heavy emotional and violent nature of the topic discussed (femicides). On the positive side, terms such as “gioia” (joy), “vera” (true), and “famiglia” (family) are present, but their frequencies are considerably lower than the negative words, highlighting a general negative emotional tone in the comments.

The second plot breaks down the top 5 words by specific emotions according to the NRC categories (e.g., anger, fear, sadness). As expected, “sadness” and “fear” categories contain highly frequent words, including “dolore”, “morte”, and “assassino”, aligning with the tragic nature of the femicide cases. The “anger” category includes “omicidio” (homicide) and “assassino”, pointing to a strong emotional response from commenters. The “joy”, “anticipation”, and “trust” categories are much less populated, reinforcing the overall dominance of negative emotional expressions in the dataset.

Overall, the plots suggest that the public’s online reactions to the femicide cases are predominantly characterized by negative emotions such as sadness, anger, and fear.

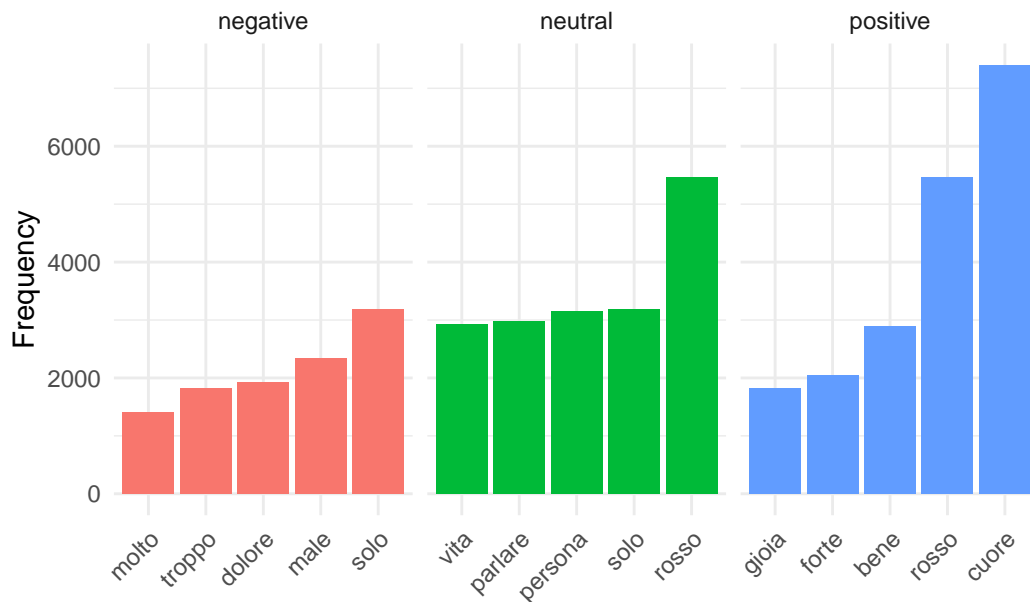


Figure 8: Word Frequencies based on the Italian Sentiment Lexicon

In the Italian Sentiment Lexicon, within the negative sentiment group, terms such as “dolore” (pain), “male” (evil), and “solo” (alone) are dominant, though their frequencies appear lower compared to the negative words highlighted in the NRC-based plot. The neutral sentiment section features common words like “vita” (life) and “persona” (person). In the positive sentiment section, words such as “cuore” (heart), “bene” (good), and “gioia” (joy) appear more frequently than the positive terms identified by the NRC lexicon. It is important to highlight that these positive words originate from the translated emoji descriptions. For example, “heart” does not necessarily convey a strictly positive sentiment, but may instead reflect a virtual expression of empathy toward the situation.

Compared to the NRC lexicon plot, the Italian Sentiment Lexicon displays a more balanced distribution, with stronger representation of both neutral and positive words. While the NRC lexicon concentrated on specific emotional categories such as anger, fear, and sadness, the Italian lexicon provides a broader polarity perspective (negative, neutral, positive). This results in a less skewed

output and highlights additional positive and neutral terms that the NRC lexicon might have under-represented. For example, words like “cuore” and “rosso” are more prominent in this classification but were less visible in the NRC analysis, reflecting differences in how each lexicon captures figurative or symbolic language. Overall, this comparison suggests that the Italian Sentiment Lexicon offers a complementary perspective, capturing not only the negativity associated with the topic but also the more neutral and positive nuances present in the comments.

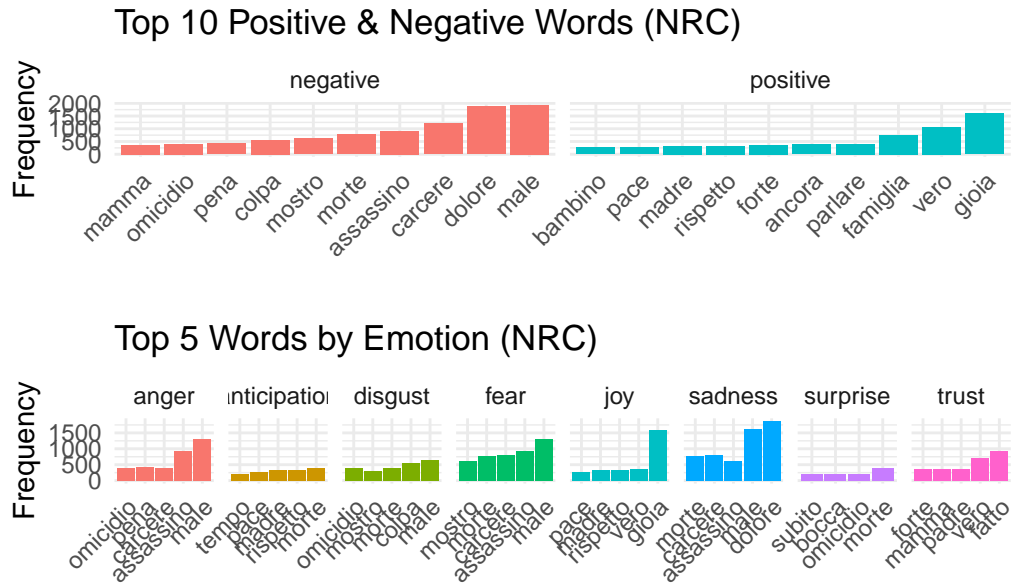


Figure 9: Sentiments expressed by Male Users

Moving on with the analysis comparing male and female users, a clear trend emerges. Negative terms dominate the chart, with “male” (evil) and “dolore” (pain) being the most frequent, followed by words such as “carcere” (prison), “assassino” (murderer), and “morte” (death). These terms highlight a strong negative emotional reaction, which is expected given the severity of the femicide cases discussed. In contrast, positive terms like “gioia” (joy), “vera” (true), and “famiglia” (family) appear less frequently, suggesting that male users’ comments tend to focus more on the tragic and violent aspects rather than on hopeful or empathetic expressions.

Emotions like sadness, fear, and anger are prominent. In particular, the sadness category includes frequent words such as “dolore”, “morte”, and “assassino”, while fear and anger include terms like “male”, “carcere”, and “omicidio” (homicide). The presence of words under surprise and trust is much more limited, and joy appears mostly with lower frequencies.

Overall, male users’ comments seem predominantly characterized by negative emotions, especially sadness and anger. This pattern suggests a discourse focused on the gravity and brutality of the events, with less space given to expressions of compassion or positivity.

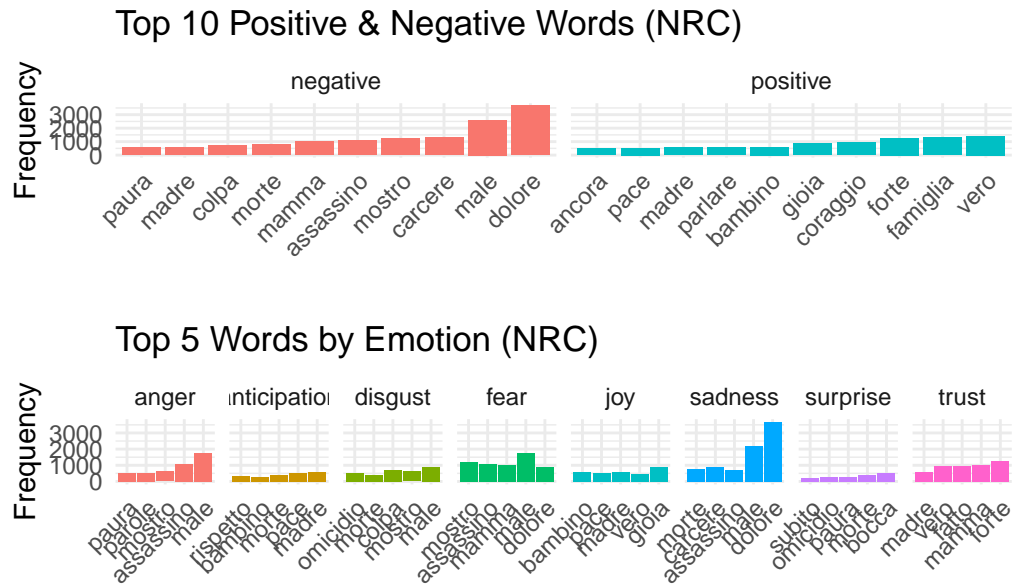


Figure 10: Sentiments expressed by Female Users

As with male users, negative words dominate, with “dolore” (pain) and “male” (evil) being the most frequent, followed by “carcere” (prison), “mostro” (monster), and “assassino” (murderer). However, compared to male users, female commenters seem to use “dolore” at a significantly higher frequency. Additionally, words like “paura” (fear) and “mamma” (mom) emerge strongly here, which were less prominent or absent among male users. This could suggest that female users are expressing more emotional and empathetic reactions in relation to the victims and the broader context of violence.

In terms of positive sentiment, words like “gioia” (joy), “famiglia” (family), “coraggio” (courage), and “vero” (true) appear. Compared to male users, female commenters mention “coraggio” (courage) more frequently, possibly reflecting a tendency to highlight resilience and support, even when discussing tragic events.

As with male users, sadness and fear are dominant, but sadness appears even stronger in this case, as shown by the prominence of “dolore”, “morte” (death), and “assassino” under this emotion. Interestingly, the fear and anger categories also show high frequencies, but with the addition of “paura”(fear) under both fear and anger. Female users appear to engage more with words connected to fear and parental empathy, as seen with “mamma” in the fear category.

When compared to male users, female commenters’ word choices appear to convey not only anger and sadness but also a notable focus on fear, protection (e.g., “mamma”, “bambino”), and expressions of courage and empathy. This may reflect gendered differences in emotional framing, where women are more likely to emphasize emotional and relational aspects in their reactions.

In summary, both male and female users display strong negative sentiment, but female users seem to place additional emphasis on fear, parental themes, and empathy, whereas male users’ com-

ments were more centered on anger and justice-oriented terms like “carcere”(prison) and “assassino”(murderer).

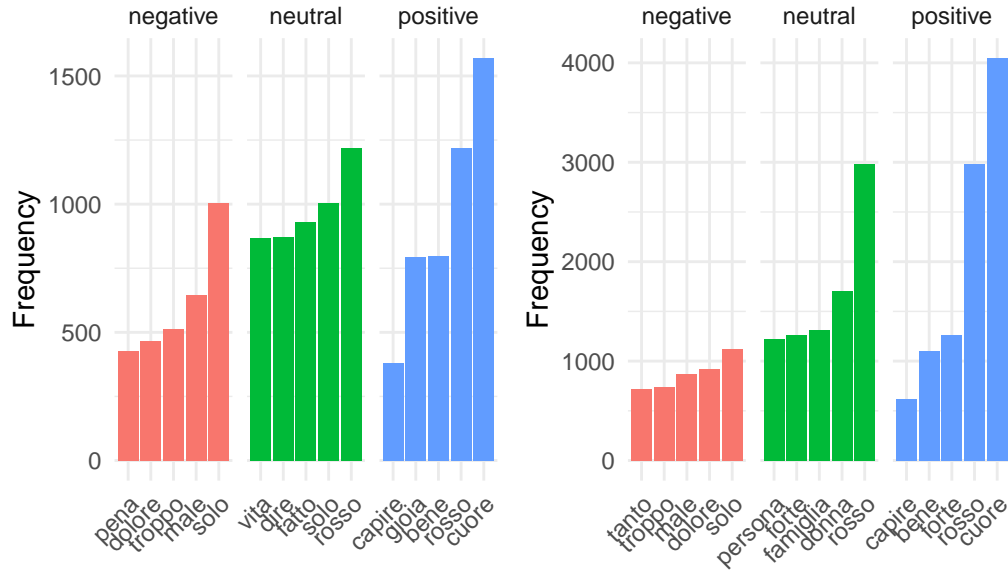


Figure 11: Sentiment between Male Users and Female Users using the Italian Sentiment Lexicon

In the male users’ plot (left), the negative sentiment section is led by terms such as “solo” (alone), “male” (evil), and “dolore” (pain), suggesting a focus on feelings of isolation, suffering, and moral judgment. In the positive sentiment section, male users mention “cuore” (heart), “bene” (good), and “gioia” (joy), but with lower frequencies than in the female panel.

In contrast, the female users’ plot (right) presents a stronger presence of positive and neutral terms. While “dolore” and “male” still dominate the negative sentiment section, their relative frequencies are lower compared to the male users. The neutral section includes emotionally significant words such as “famiglia” (family) and “mamma” (mom), suggesting a focus on relational and protective language more typical in female discourse. Positive sentiment is more prominent among female users, especially with words like “cuore”(heart), and “forte” (strong), which appear with higher frequencies than in the male group.

Overall, this comparison shows that male users focus slightly more on negative sentiments related to suffering and moral judgment, while female users show a stronger tendency toward expressing relational and emotional values, particularly through neutral and positive terms such as “famiglia,” “mamma,” “cuore,” and “forte.” This difference may indicate a more empathetic and nurturing tone in female comments, whereas male comments may lean more toward expressions of pain and condemnation.

In this sense, both lexicons highlight the same overarching trend between male and female users.

The analysis now shifts to exploring the sentiment patterns of the most active and less active users.

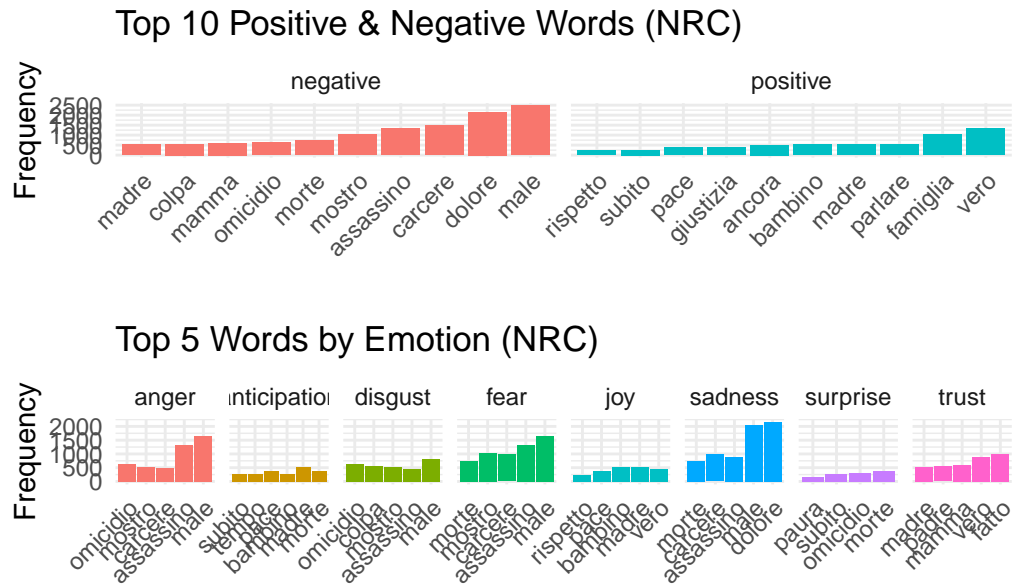


Figure 12: Sentiments expressed by Most Active Users

In the top chart, negative terms dominate, led by “male” (evil) and “dolore” (pain), followed by “carcere” (prison), “assassino” (murderer), and “morte” (death). Positive words like “famiglia” (family), “vero” (true), and “bambino” (child) appear but with lower frequencies, suggesting that highly active users focus more on the tragic and violent aspects of the events. In the bottom chart, the most frequent emotions are sadness and fear, as shown by words such as “dolore” (pain), “morte” (death), and “assassino” (murderer). The anger and disgust categories are also relevant, with terms like “omicidio” (homicide) and “mostro” (monster) appearing. Positive emotions such as “joy” and “trust” include fewer and less frequent words, such as “rispetto” (respect) and “giustizia” (justice). Overall, the most active users’ comments are heavily centered on negative emotions, especially sadness and fear, with limited expression of positive or empathetic sentiment.

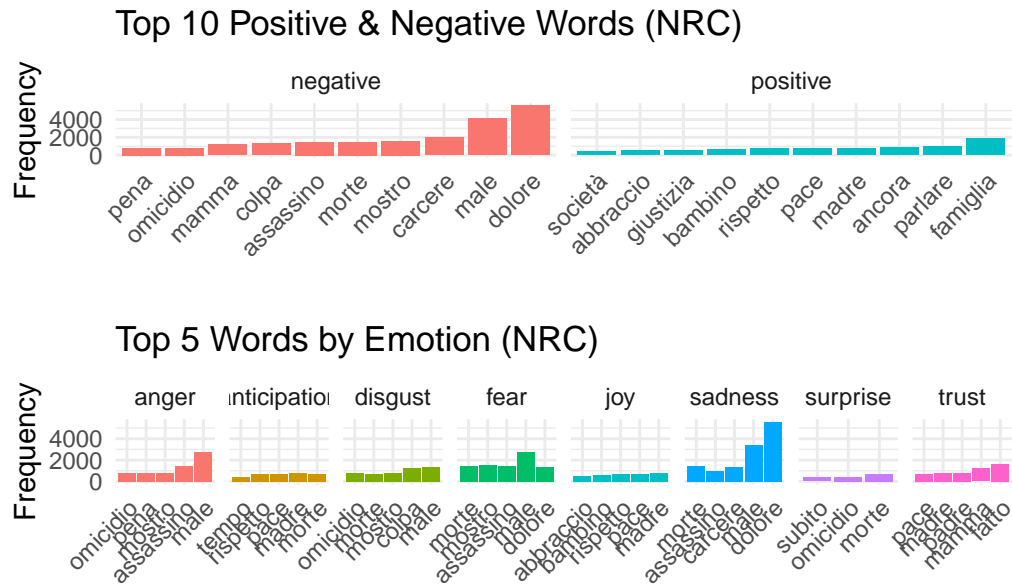


Figure 13: Sentiments expressed by Less Active Users

In the top chart, negative terms are predominant, with “dolore” (pain) and “male” (evil) being the most frequent, followed by “carcere” (prison), “mostro” (monster), and “morte” (death). Compared to the most active users, there is a higher frequency of words like “solitudine” (loneliness, appearing as “sola” = alone) and “abbandono” (abandonment, reflected in “abbastanza” = quite/enough as a stemmed form of other words). On the positive side, words such as “famiglia” (family), “rispetto” (respect), and “bambino” (child) appear but with lower frequencies compared to negative terms. In the bottom chart, negative emotions such as sadness, fear, and anger dominate. The sadness category shows very high frequencies for “dolore” (pain), “morte” (death), and “assassino” (murderer). Fear-related terms, such as “carcere” (prison) and “male” (evil), are also present. Overall, less active users express a heavier emotional load, especially through sadness and fear, but also introduce more isolated mentions of justice-related words like “giustizia” (justice) and “pace” (peace) within positive sentiments, suggesting occasional appeals to resolution or empathy.

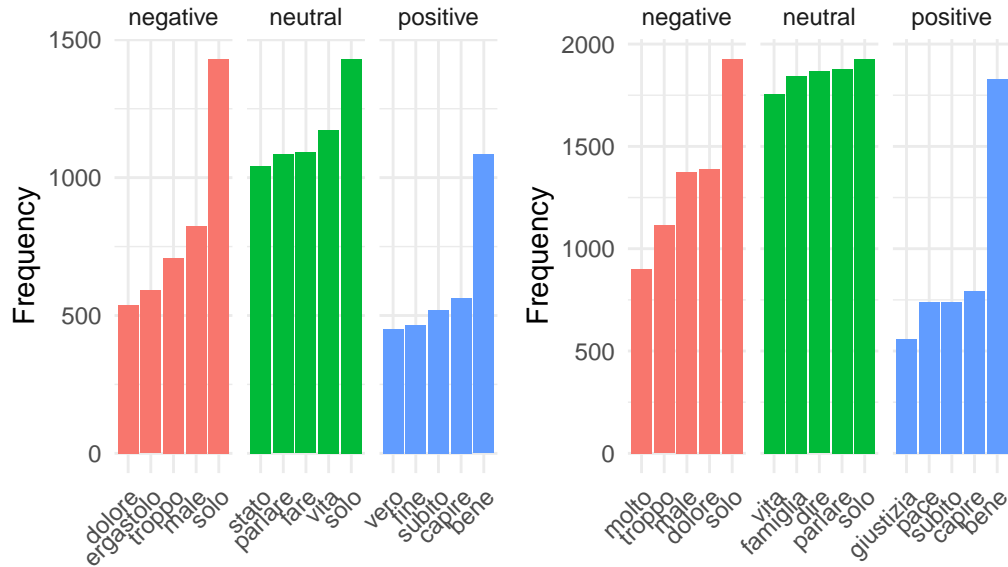


Figure 14: Sentiment between Most Active Users and Less Active Users using the Italian Sentiment Lexicon

In the left panel, most active users show a predominance of negative terms such as “solo” (alone), “male” (evil), “troppo” (too much), and “dolore” (pain), reflecting feelings of isolation and emotional distress. The neutral sentiment includes generic and connective terms like “parlare” (to speak), “fare” (to do), and “vita” (life), while positive sentiment is represented by words like “bene” (good), “capire” (to understand), and “subito” (immediately). However, the frequencies of positive words remain lower compared to the negative and neutral ones, suggesting that highly active users tend to repeat emotionally heavy or negative expressions more often.

In the right panel, less active users exhibit a stronger focus on negative terms such as “solo” (alone), “dolore” (pain), and “male” (evil), but with higher frequencies compared to the most active group. The neutral sentiment is dominated by words like “vita” (life), “famiglia” (family), and “diritto” (right), which introduce more relational and societal concepts compared to the most active users. In the positive sentiment group, less active users more frequently mention “giustizia” (justice), “bene” (good), and “capire” (to understand), suggesting that while their comments may focus on negative events, they also express hope for resolution or moral reflection.

Overall, while both groups emphasize negative sentiment, less active users seem to incorporate slightly more references to relational values and justice, while the most active users concentrate on more repetitive and emotionally charged terms like “solo” and “dolore”. This confirms the recurrence of similar patterns between the two lexicons, even when analyzing these subgroups.

Positive and Negative Sentences

This section provides an analysis of the most positive and most negative sentences across the entire dataset and within each subgroup. The analysis is conducted using both lexicons. In this case, two versions of the text are used: one where emojis are replaced with their descriptions, and one where emojis are removed entirely. Separate analyses are then performed on each corresponding corpus.

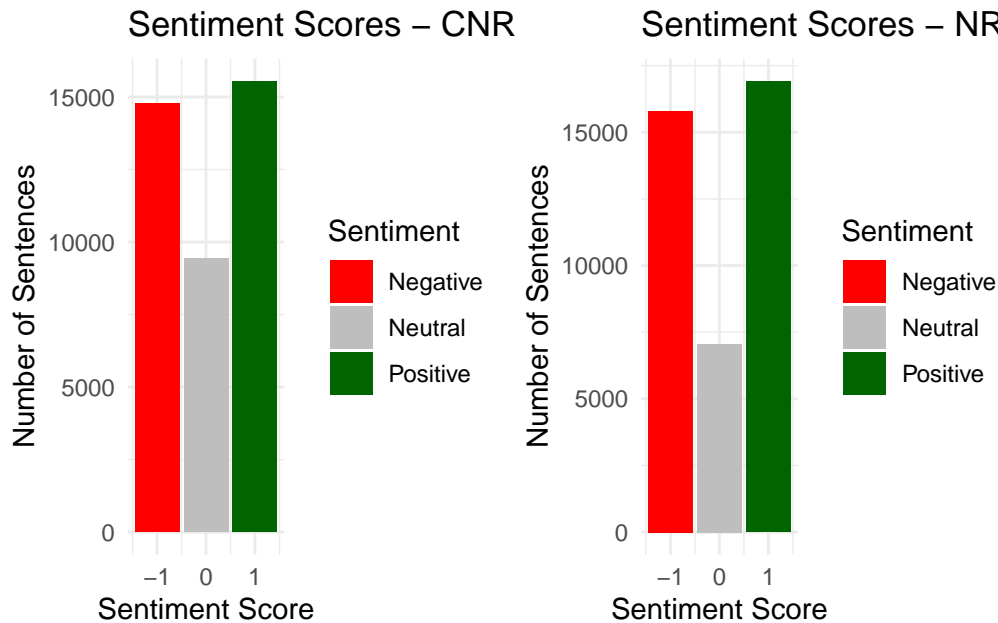


Figure 15: Sentences Sentiment without Emoji

The two barplots present the distribution of sentiment scores in sentences where emojis have been erased, using two different lexicons: CNR on the left and NRC on the right. In both cases, the majority of sentences are classified as either positive or negative, with neutral sentences being much fewer. This suggests that even without emojis, the textual content is rich enough to convey clear sentiment. Comparing the two lexicons, the CNR lexicon results in a slightly higher number of positive sentences compared to negative ones, indicating a more positive skew. The NRC lexicon, on the other hand, shows a more balanced distribution between positive and negative sentences, though positive sentences are still slightly more frequent. In both cases, the number of neutral sentences remains consistently lower. Overall, the choice of lexicon influences how sentiment is captured after removing emojis, with CNR leaning more towards positivity and NRC showing a more even spread between positive and negative classifications.

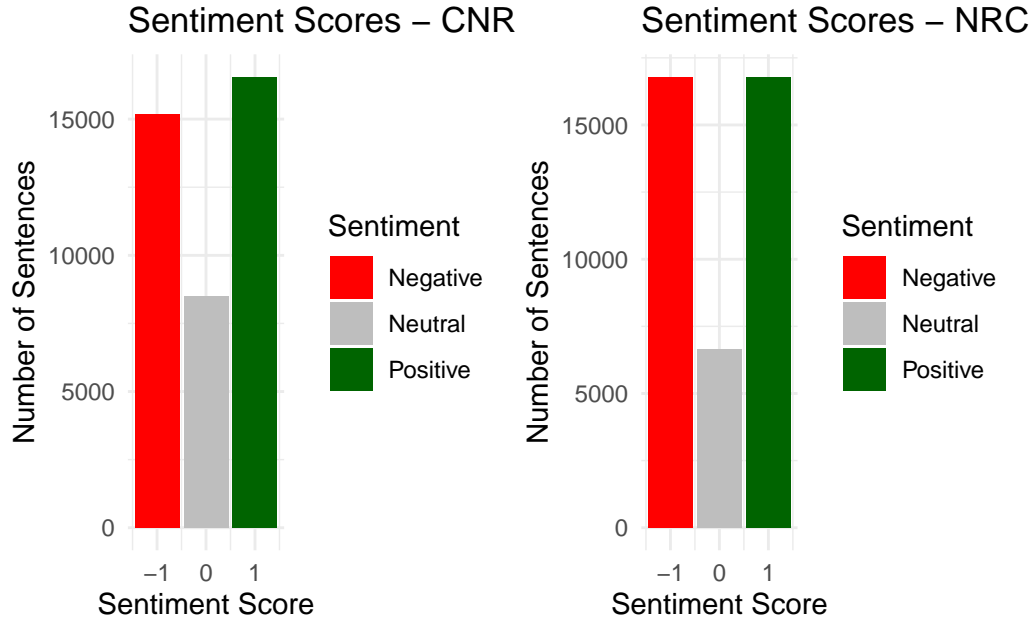


Figure 16: Sentences Sentiment with Emoji

In this case, where the emojis have been replaced with their descriptions, the overall distribution of sentiment scores remains polarized, with a majority of sentences classified as either positive or negative and fewer neutral sentences. This suggests that the emoji descriptions continue to contribute meaningfully to the sentiment classification. For the CNR lexicon, there is a slight predominance of positive sentences compared to negative ones, mirroring the previous trend seen when emojis were erased. The NRC lexicon shows a very balanced distribution between positive and negative sentences, with both categories occurring almost equally. The replacement of emojis with their textual descriptions helps maintain the sentiment-bearing function of the original emojis, reinforcing polarity and reducing the number of neutral classifications. The choice of lexicon continues to influence the results, with CNR showing a marginal positive skew and NRC maintaining a more balanced classification between positive and negative sentiments.

Since these results indicate that emojis play a significant role in polarizing sentence sentiment, the subsequent analyses on the other subgroups are conducted exclusively using the text in which emojis have been replaced with their descriptions.

Starting with the analysis of the sentiment of the sentences between female and male users:

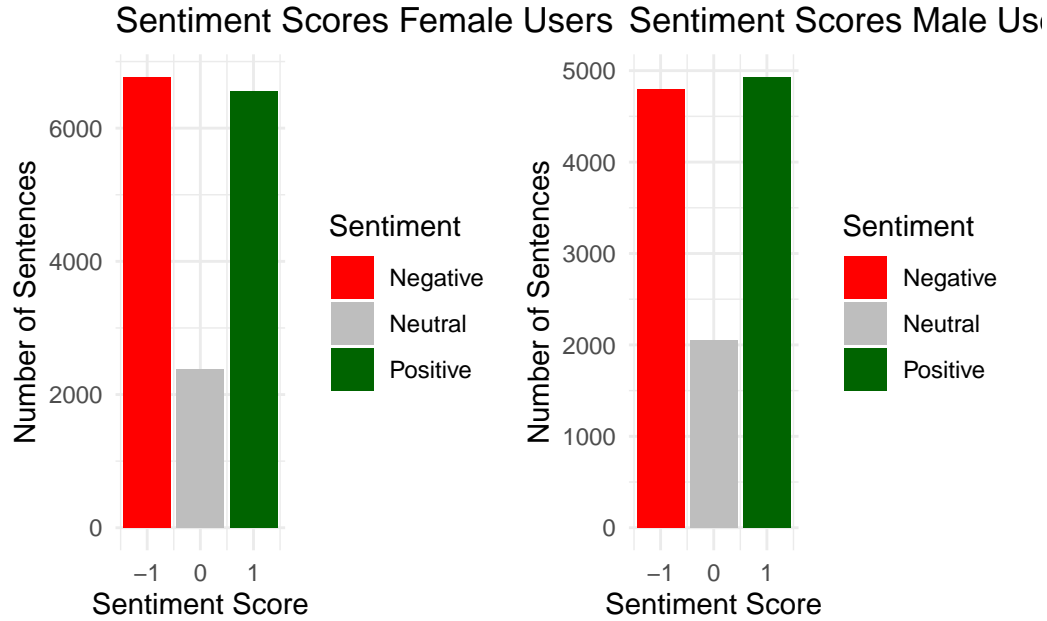


Figure 17: Sentences Sentiment between Genders with NRC

For female users (left plot), there is a slightly higher number of negative sentences compared to positive ones, though both categories are relatively balanced. Neutral sentences are notably fewer, indicating that most sentences from female users express clear sentiment polarity. For male users (right plot), the trend is similar, with negative and positive sentences occurring in almost equal proportions. However, in this case, positive sentences slightly outnumber negative ones. Neutral sentences again appear in smaller quantities. Overall, both groups show a low occurrence of neutral sentences, suggesting that both male and female users tend to write sentences with a distinct sentiment. However, female users exhibit a slight tendency toward negativity, while male users show a slight inclination toward positivity according to the NRC lexicon.

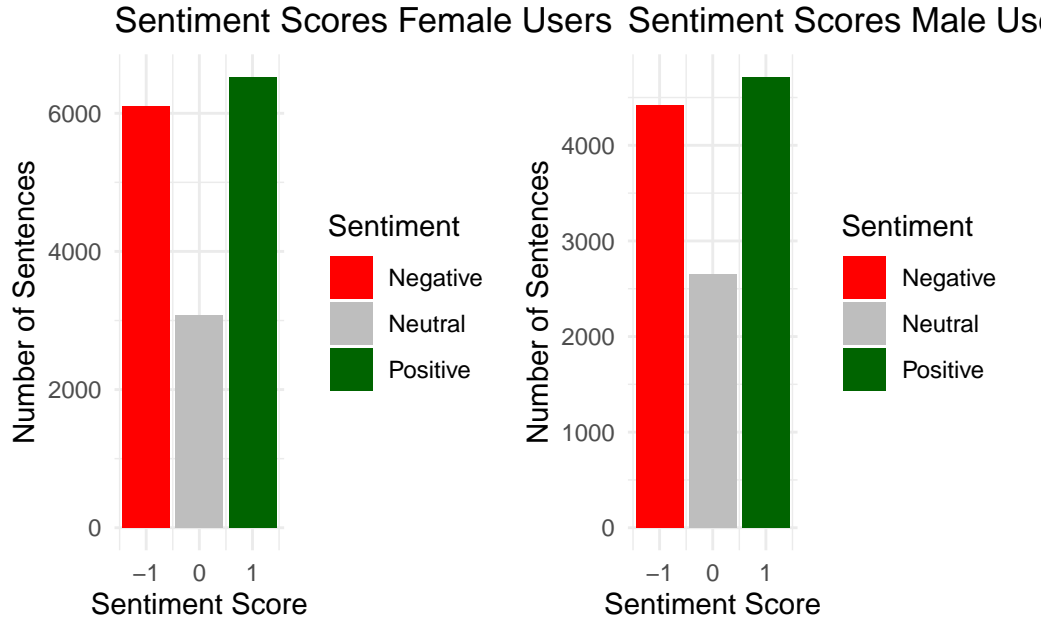


Figure 18: Sentences Sentiment between Genders with the Italian Sentiment Lexicon

For female users (left plot), positive sentences slightly outnumber negative ones, suggesting a moderate tendency toward positive sentiment. Neutral sentences are less frequent, indicating that female users tend to express clear emotional polarity. For male users (right plot), the pattern is similar, with positive sentences being noticeably more frequent than negative ones. The gap between positive and negative sentences is larger compared to female users, showing a stronger inclination toward positivity in male users. Neutral sentences are again less common. Overall, when using the CNR lexicon, both female and male users predominantly produce sentences with clear sentiment, with male users showing a slightly stronger positive skew compared to female users.

Comparing this plot to the previous one where the NRC lexicon was used, it is evident that the two dictionaries behave more differently when it comes to sentiment scores than they do when analyzing word-level sentiment. While both lexicons show a general polarity in sentence-level sentiment, the differences between them are more pronounced at the sentence level. Specifically, the NRC lexicon tends to yield a more balanced distribution between positive and negative sentences, while the CNR lexicon shows a clearer tendency toward positivity, especially among male users. This suggests that the choice of lexicon has a greater impact on sentence-level sentiment analysis than on the classification of individual words.

Continuing with the analysis comparing most active and less active users:

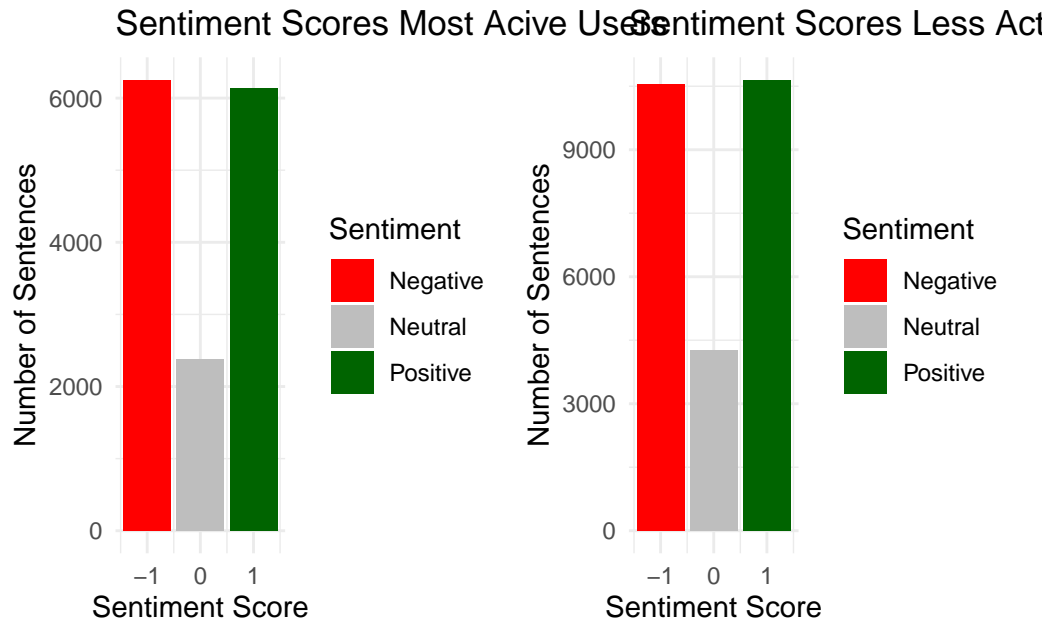


Figure 19: Sentences Sentiment between Genders with NRC

For the most active users (left plot), negative and positive sentences occur in nearly equal proportions, with a slight predominance of negative sentences. Neutral sentences are considerably fewer, suggesting that active users tend to express polarized sentiments. For the less active users (right plot), the pattern is similar but with an even more balanced distribution between negative and positive sentences. Positive sentences slightly outnumber the negative ones. Neutral sentences remain the least common category but are slightly more frequent compared to the most active users.

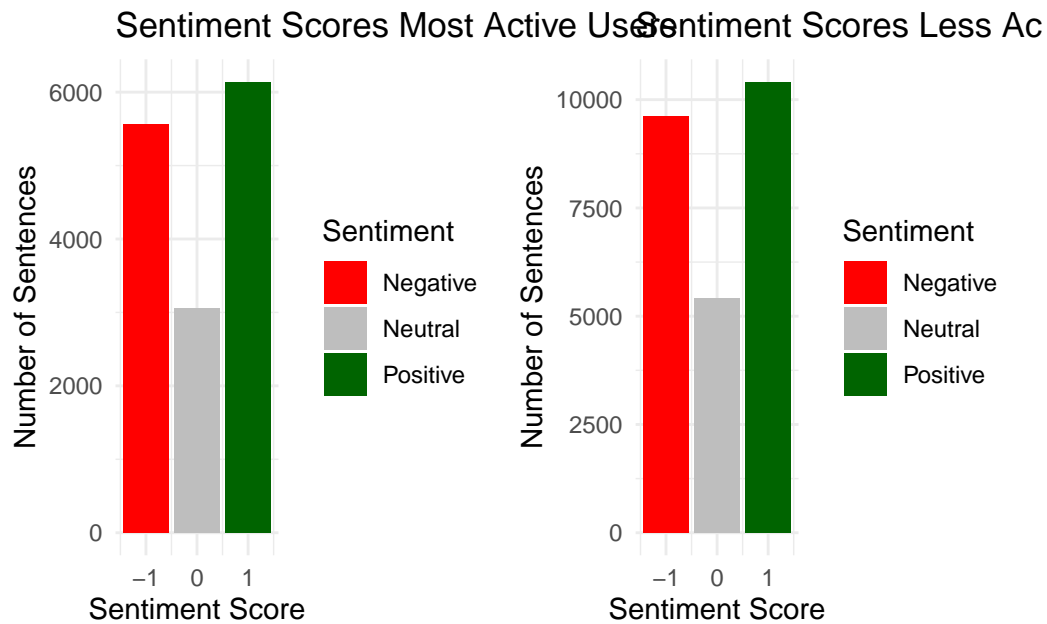


Figure 20: Sentences Sentiment between Genders with the Italian Sentiment Lexicon

For the most active users, positive sentences slightly outnumber negative ones, indicating a mild tendency toward positivity. Neutral sentences are less frequent, suggesting that highly active users tend to write sentences with clear emotional polarity. For the less active users, the trend is even more pronounced. Positive sentences are noticeably more frequent than negative ones, and neutral sentences remain the least common. Compared to the most active users, less active users display a stronger positive skew.

Once again, the two dictionaries do not show a strong correspondence in sentence-level sentiment scores. While both the CNR and NRC lexicons highlight polarized sentiment, the way they classify sentences differs notably. This reinforces the idea that the choice of dictionary has a stronger impact on sentence-level sentiment results than on word-level sentiment, leading to diverging interpretations depending on which lexicon is applied.

BI-GRAMS

The analysis of bigrams, or pairs of adjacent words, provides valuable insight into the relational structure and thematic patterns within the dataset. By focusing on word pairs instead of isolated tokens, it is possible to better capture contextual meaning and detect recurring expressions that convey sentiment, judgment, or actions. In this work, bigrams are extracted after applying stemming to reduce lexical variability, followed by the reconstruction of original terms for interpretation. This approach allows for the identification of key expressions that shape the discourse, highlighting how users articulate emotions, moral stances, and calls to action in the context of femicide-related discussions.

The first bigram analysis is performed on the full dataset, using the corpus where emoji descriptions have been removed.

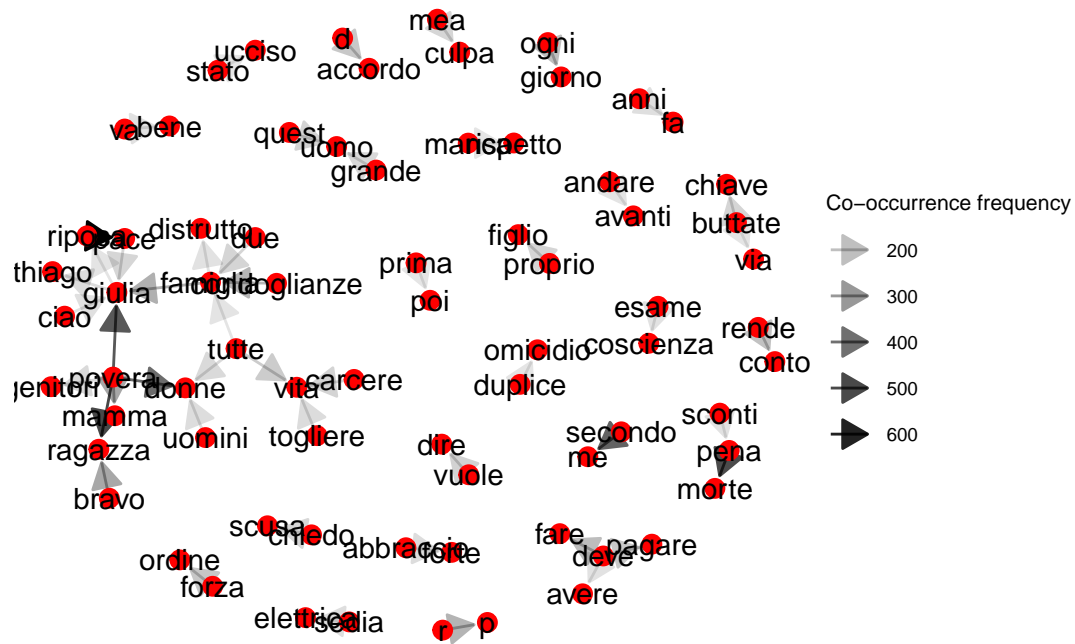


Figure 21: Bigram Network on the Full Dataset

The graph highlights key themes and patterns within this discourse. Some clusters focus on expressions of empathy and condolence, such as “condoglianze famiglia” (condolences family) and “abbraccio forte” (strong hug), suggesting that users often express solidarity and emotional support. Another cluster revolves around judgment and guilt, with bigrams like “colpa mea” (my fault), “coscienza esame” (conscience exam), and “pena morte” (death penalty), indicating a focus on moral responsibility and justice. There is also a noticeable emphasis on the victim and family, with combinations like “povera mamma” (poor mother), “ragazza povera” (poor girl), and “genitori famiglia” (parents family), reflecting emotional responses centered around the victims and their close ones. Lastly, verbs such as “deve essere” (must be), “può fare” (can do), and “buttate via” (throw away) indicate calls to action or reflection, pointing to public discourse about what should or could be done.

This second plot compares the bigram networks of male and female users to identify potential differences in patterns, themes, and language use between the two groups.

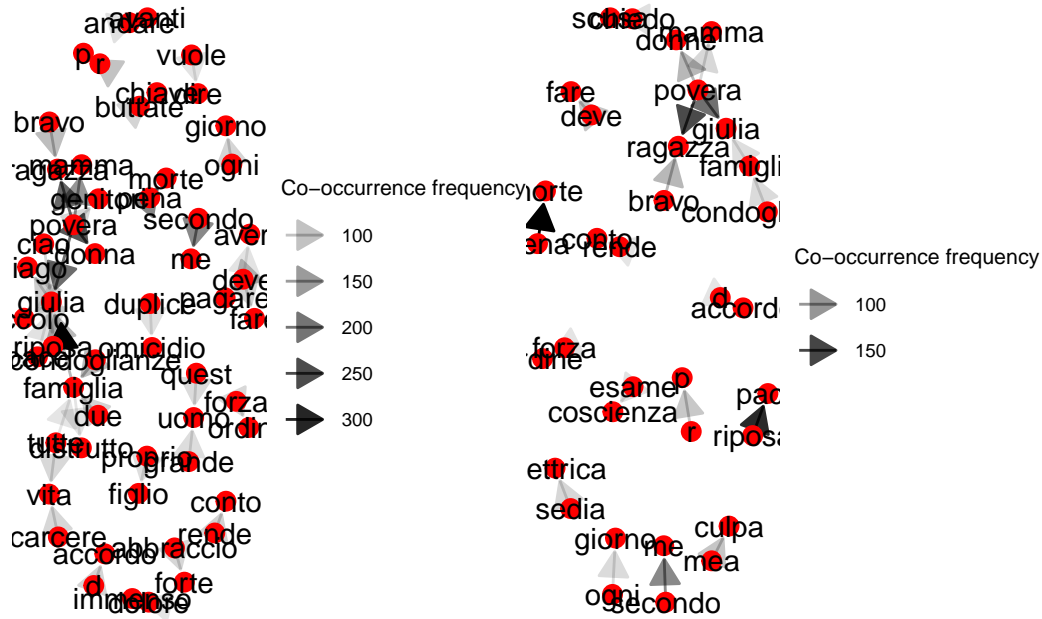


Figure 22: Bigram Network for Female and Male Users

In both networks, emotional and socially charged terms dominate the discourse. Female users (left plot) often produce bigrams related to empathy and mourning, such as “povera ragazza” (poor girl), “abbraccio forte” (strong hug), and “condoglianze famiglia” (condolences family), indicating a focus on emotional support and compassion. There is also a presence of justice-related terms like “pena morte” (death penalty) and “omicidio duplice” (double homicide), reflecting concerns about punishment and accountability. Male users (right plot) show a similar tendency toward expressions of sympathy and sorrow with bigrams like “povera mamma” (poor mother), “riposa pace” (rest in peace), and “famiglia condoglianze” (family condolences). However, compared to female users, male comments display fewer references to punitive terms and more toward traditional condolences and reactions such as “bravo ragazza” (good girl) or “giulia famiglia” (Giulia family), which are more personalized. Additionally, male users’ bigram network appears more fragmented, suggesting slightly less lexical variety in co-occurrence patterns compared to female users. Female users tend to have a denser and more connected network of expressions involving emotions, actions, and moral judgments.

Lastly, this plot compares the bigram networks of the most active users with those of the less active users.

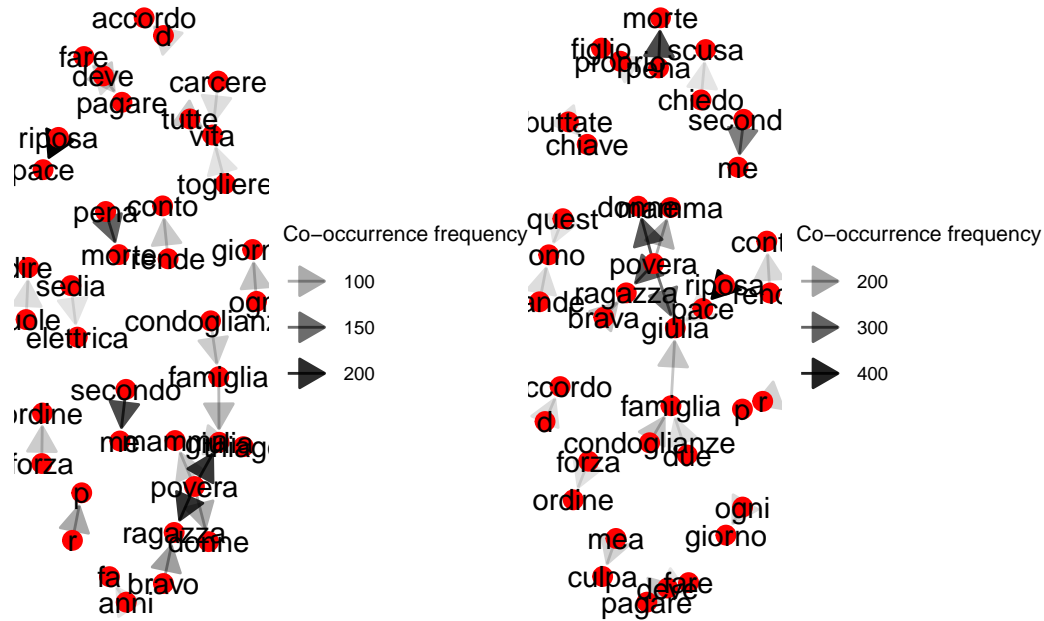


Figure 23: Bigram Network on Most Active and Less Active Users

For the most active users (left plot), the network highlights terms related to both condolence and justice. There are expressions of sympathy such as “condoglianze famiglia” (condolences family) and “riposa pace” (rest in peace), but also stronger justice-oriented bigrams like “pena morte” (death penalty), “sedia elettrica” (electric chair), and “togliere vita” (take life). The presence of these terms suggests that active users engage in discussions not only to show grief but also to express anger and seek justice or punishment. Less active users (right plot) also show strong emotional involvement, with frequent bigrams such as “riposa pace” (rest in peace), “famiglia condoglianze” (family condolences), and “povera ragazza” (poor girl). However, compared to active users, their network appears to lean more toward mourning and sympathy rather than calls for punitive justice. The justice-oriented terms are present but less prominent, with fewer edges linked to expressions like “pena morte”. Additionally, the network of less active users appears slightly denser, with key terms like “giulia”, “mamma”, “povera”, and “famiglia” playing a central role, reflecting a stronger emotional focus on the victims and their families. In summary, both groups display emotional reactions, but most active users are more likely to mix sympathy with discussions of punishment and justice, whereas less active users show a tendency to focus more on sorrow, condolences, and empathy for the victims. This suggests a difference in engagement style between the two groups: active users balance grief with societal critique, while less active users center more on emotional and familial aspects.

The Impact of Negation on Sentiment Polarity

Negation plays a crucial role in shaping the emotional tone of language, particularly in sentiment analysis. In this section, the focus is on exploring how different forms of negation influence the

polarity of words in user comments. By identifying common negation words and examining the terms that follow them, it becomes possible to assess how sentiment shifts under the presence of negation. Two lexicons, NRC and CNR, are applied separately to evaluate whether sentiment patterns remain consistent or diverge depending on the sentiment dictionary used. The following plots highlight how negated contexts contribute to either amplifying or modifying the emotional weight of words within the dataset.

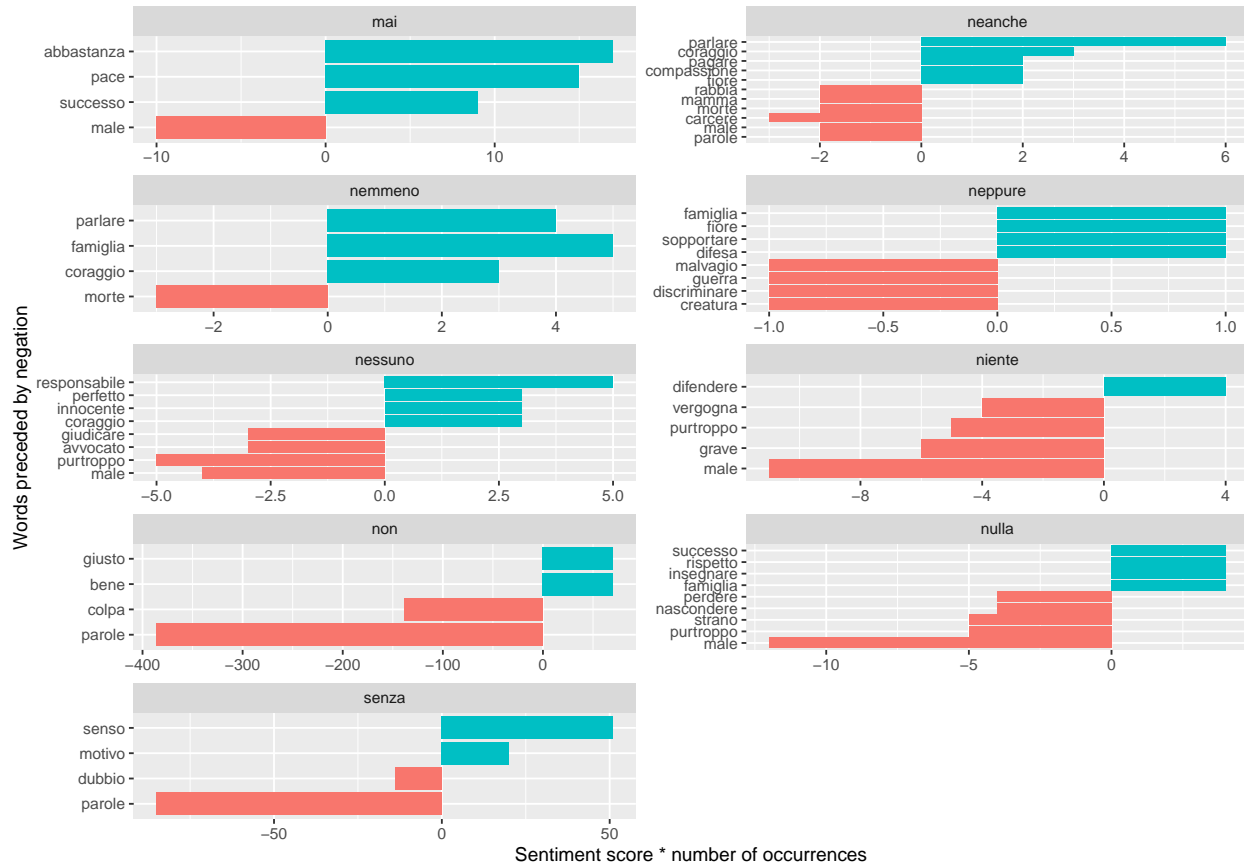


Figure 24: Positive and Negative emotions with NRC

The bars represent the product of sentiment score and frequency for each negated word, offering insight into both the intensity and direction of sentiment under negation. The analysis reveals that negations often shift the sentiment context. For instance, words like “parole” and “colpa” following non exhibit strong negative polarity, while words such as “giusto” and “bene” maintain a positive orientation even when negated. The word “male” consistently carries a negative sentiment across various negators like nessuno, niente, and nulla, reinforcing its strong negative connotation. Interestingly, some positive or neutral words like “famiglia” and “coraggio” appear after negators such as mai and nemmeno, suggesting more complex or potentially ironic sentiment patterns. Additionally, “senza” (without) enhances negative framing, particularly in combinations like “senza parole” (speechless) and “senza dubbio” (without doubt). This plot underscores how negation, combined with the NRC lexicon, influences the emotional tone of comments by amplifying negativity or creating nuanced sentiment shifts, especially in sensitive discussions like those surrounding femicides.

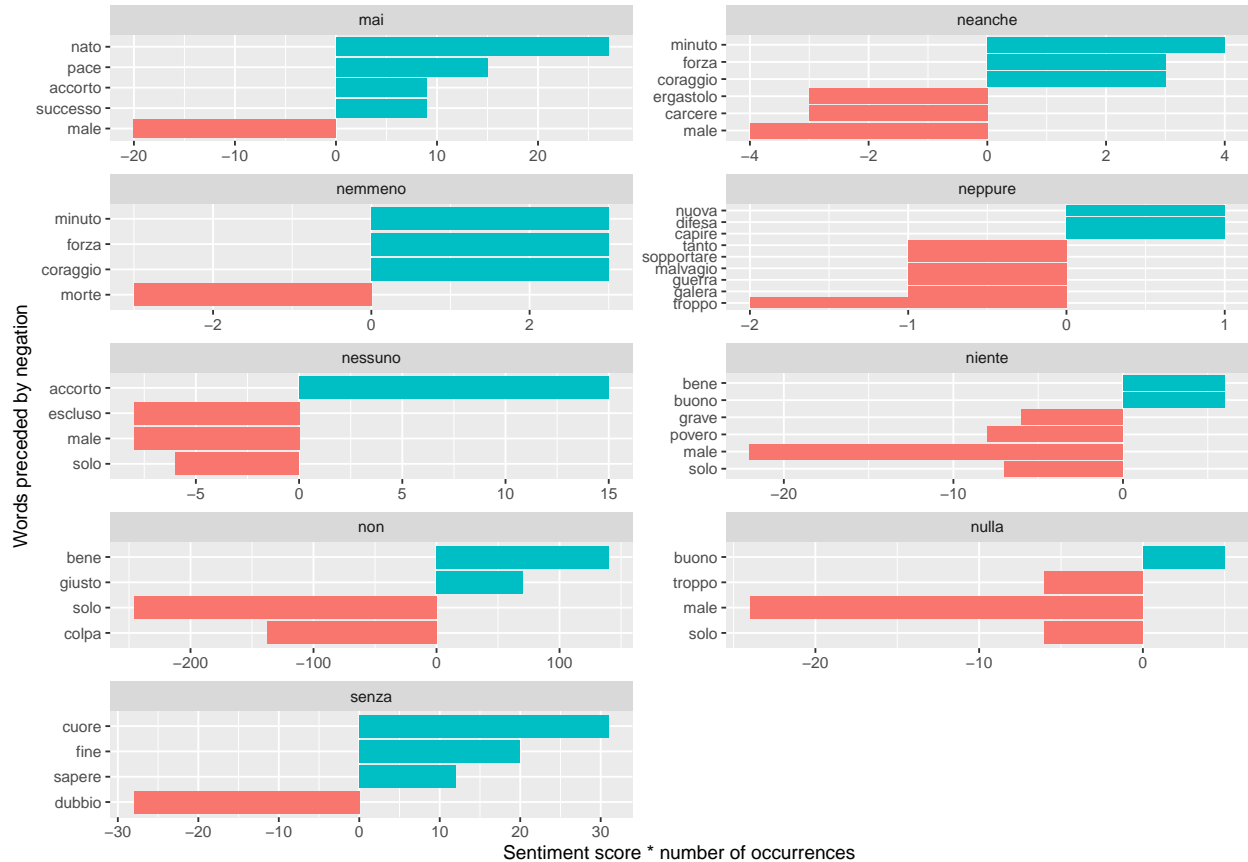


Figure 25: Positive and Negative emotions with CNR

Compared to the NRC version, the CNR lexicon yields stronger negative polarities, particularly after negations such as *niente*, *nulla*, and *non*. For example, words like “solo”, “male”, and “colpa” consistently carry a heavier negative sentiment load, especially under *non*, *niente*, and *nulla*. The strong negativity reflects how the CNR lexicon emphasizes harsh or judgmental tones in these contexts. On the positive side, “bene”, “giusto”, and “buono” retain their positivity after negation but with less intensity than in the NRC-based plot, suggesting that the CNR lexicon tends to mitigate positive polarities when combined with negators. Additionally, words following *senza* (without), such as “cuore” (heart) and “fine” (end), display both negative and positive sentiment, reinforcing that negation can generate mixed emotional tones, especially in sensitive discussions like those related to femicides. Overall, with the CNR lexicon, the impact of negation leans more heavily towards reinforcing negative sentiment, as compared to the more balanced effect observed with the NRC lexicon. This highlights how different lexicons can shape the perceived sentiment dynamics when negation is present.

Discussion and Limitations

While the gender assignment process enables to conduct the analysis despite the lack of self-reported gender data, several limitations should be acknowledged. First, usernames are inherently unreliable

indicators of identity. They may reflect nicknames, fictional characters, brand names, or other non-human entities (e.g., “MusicLover1990” or “Tech_Company”). This could lead to erroneous gender assignments, as the method assumes that certain name patterns embedded in usernames relate to actual individuals. Second, even when usernames include a valid first name, this may not correspond to the gender identity of the commenter. The method does not account for cases where individuals use names that do not align with their gender identity, nor does it consider non-binary or gender-diverse identities. For simplicity, the system applies a binary gender model (male/female) and resorts to an “unknown” category when no reliable assignment is possible. Third, both the name-matching step and the LLM rely on probabilistic and external datasets, which may embed their own biases (e.g., over-representation of certain names, traditional gender associations in datasets, or model biases from the LLM). These biases can impact the gender classification and affect the analysis. Finally, this method assumes that one username corresponds to one individual, ignoring cases where accounts may be shared (e.g., organizational accounts) or where one person operates multiple accounts. Despite these limitations, introducing the “unknown” category reduces the risk of misclassification by acknowledging uncertainty. Nonetheless, the presence of the “unknown” category must be considered when interpreting the results, as it could influence aggregate gender-related trends in the data. These ethical and methodological concerns are important to bear in mind when evaluating the validity and generalizability of the findings, and further studies could benefit from alternative approaches.

A part from the limitations regarding the gender assignment procedures, some more arised in the analysis:

One of the key limitations identified in this work relates to the dictionaries used for sentiment analysis. While lexicons like NRC and CNR are widely employed in sentiment research, they proved to be unsatisfactory in fully capturing the nuances of the language within this dataset. Certain sentiment attributions appeared difficult to justify, even for native speakers, suggesting that these resources may oversimplify or misrepresent the emotional tone of specific terms in context.

Another important consideration is that the sentiment evaluation was conducted primarily at the word level, without a broader contextual understanding of full sentences or comments. This approach limits the ability to capture the complexity and subtlety of the discourse, where negation, irony, and multi-word expressions could alter the intended sentiment. A sentiment analysis framework that considers entire comments, rather than isolated tokens, would likely provide a more accurate depiction of the users’ emotional expressions.

Additionally, this analysis could benefit from the integration of lemmatization. Unlike stemming, which mechanically reduces words to their root forms, lemmatization relies on a dictionary to return the correct base or dictionary form of a word. This makes lemmatization more suitable for preserving meaning and reducing lexical variability, as it groups together words that share the same lemma but differ in inflection. Incorporating lemmatization would likely improve the consistency of sentiment scoring and enhance the interpretability of the bigram networks.

Lastly, the study did not account for the presence of multiple identical emojis within a single comment. By treating emojis as binary elements (present or erased) without quantifying their frequency, the analysis may have introduced a bias. Repeated emojis could potentially amplify the intended

sentiment of a comment, and neglecting this dimension may have led to an underestimation or distortion of their emotional impact.

Conclusion

This research has provided valuable insights into how male and female users, as well as users with varying levels of activity, engage with discussions surrounding femicides on YouTube. By analyzing comments on the femicides of Giulia Cecchettin and Giulia Tramontano, this study demonstrates that gender plays a significant role in shaping the emotional and discursive tone of online reactions.

The results confirm that both male and female users tend to express predominantly negative sentiments, as expected given the tragic subject matter. However, notable differences emerged. Female users exhibited a stronger focus on empathy, fear, and relational themes, as shown by the frequent use of terms such as “mamma,” “famiglia,” and “coraggio.” Male users, while also expressing grief and anger, were more likely to emphasize punitive or justice-related expressions, including “carcere,” “pena,” and “assassino.” Similar patterns were observed when comparing the most active users to less active users, where highly active users displayed a greater focus on justice and condemnation, while less active users were more inclined to express sorrow and empathy.

Furthermore, the analysis of negation revealed that negation structures amplify negative sentiment, with both lexicons (NRC and CNR) showing that words preceded by negators such as *non* or *senza* tend to lean more heavily toward negative polarity. However, discrepancies between the lexicons highlight how methodological choices affect sentiment interpretation. The NRC lexicon showed a more balanced polarity distribution, while the CNR lexicon amplified negative scores, particularly in combination with negation.

The project also revealed limitations in current sentiment lexicons, as certain sentiment scores appeared misaligned with the contextual meaning recognized by native speakers. This suggests that existing lexicons may not fully capture the complexity of language in sensitive topics such as femicide. Moreover, analyzing sentiment primarily at the word level proved limiting, as it failed to account for broader contextual or syntactic structures within full sentences or comments. A future extension should consider sentiment models capable of sentence-level analysis.

Another avenue for improvement is the use of lemmatization, which—unlike stemming—relies on dictionary-based processing to resolve inflections while preserving semantic meaning. Integrating lemmatization would likely enhance sentiment scoring and improve the interpretability of bigram networks.

Finally, a potential bias was introduced by the handling of emojis. This study replaced or erased emojis without accounting for their frequency within comments. Repeated emojis may intensify the emotional charge of a comment, and this aspect should be considered in future sentiment models.

In conclusion, this research contributes to the understanding of gendered patterns in online discourse around femicide and illustrates how methodological choices (e.g., lexicon selection, text

pre-processing) can impact findings. The study encourages further exploration of gender and emotional expression in digital spaces and suggests that integrating more context-aware tools could improve the reliability and depth of future sentiment analyses on social issues.

References

- Agenzia Spada. (n.d.). *Spot pubblicitari su La7: informazioni e costi*. Retrieved from <https://agenziaspada.com>
- Agenzia Spada. (n.d.). *Spot pubblicitari su Nove: informazioni e costi*. Retrieved from <https://agenziaspada.com>
- Ansa.it. (2023). *Giulia uccisa con 37 coltellate, 2 fatali al collo*. Retrieved from <https://ansa.it>
- AP News. (2024). *Italy holds a trial into the killing of a woman that sparked debate over femicide*. Retrieved from <https://apnews.com>
- Arxiv.org. (2024). *Gender differences in online communication: A case study of Soccer*. Retrieved from <https://arxiv.org/abs/2403.11051>
- CEUR-WS.org. (2023). *Beyond Headlines: A Corpus of Femicides News Coverage in Italian Newspapers*. Retrieved from <http://ceur-ws.org>
- Küchler, C., Stoll, A., Ziegele, M., & Naab, T.K. (2023). *Gender-Related Differences in Online Comment Sections*. Sage Journals. <https://doi.org/10.1177/19401612221137905>
- Wikipedia. (n.d.). *La7 - Wikipedia*. Retrieved from <https://it.wikipedia.org/wiki/La7>
- Thelwall, M., & Foster, D. (2021). *Male or female gender-polarized YouTube videos are less viewed*. *Journal of the Association for Information Science and Technology*.
- Wikipedia. (n.d.). *Nove (rete televisiva) - Wikipedia*. Retrieved from [https://it.wikipedia.org/wiki/Nove_\(rete televisiva\)](https://it.wikipedia.org/wiki/Nove_(rete televisiva))
- People.com. (2023). *Man Who Hid Murdered Girlfriend's Body in Garbage Bags at Ravine Sentenced to Life in Case That Shocked Italy*. Retrieved from <https://people.com>
- ResearchGate. (2022). *Femicide in Italy: "Femminicidio", moral panic and progressivist discourse*. Retrieved from <https://researchgate.net>
- Sage Journals. (2023). *Investigating gender and racial-ethnic biases in sentiment analysis of language*. Retrieved from <https://tandfonline.com>
- UN Women. (2023). *Cinque fatti essenziali da sapere sul femminicidio*. Retrieved from <https://unwomen.org>
- YouTube. (n.d.). *NOVE - Canale ufficiale su YouTube*. Retrieved from <https://youtube.com>

Appendix

```
print(conf_matrix)
```

Confusion Matrix and Statistics

Reference

Prediction	male	female	unknown
male	29	4	5
female	0	39	5
unknown	0	0	18

Overall Statistics

Accuracy : 0.86
 95% CI : (0.7763, 0.9213)
 No Information Rate : 0.43
 P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.7847

McNemar's Test P-Value : 0.002905

Statistics by Class:

	Class: male	Class: female	Class: unknown
Sensitivity	1.0000	0.9070	0.6429
Specificity	0.8732	0.9123	1.0000
Pos Pred Value	0.7632	0.8864	1.0000
Neg Pred Value	1.0000	0.9286	0.8780
Prevalence	0.2900	0.4300	0.2800
Detection Rate	0.2900	0.3900	0.1800
Detection Prevalence	0.3800	0.4400	0.1800
Balanced Accuracy	0.9366	0.9096	0.8214

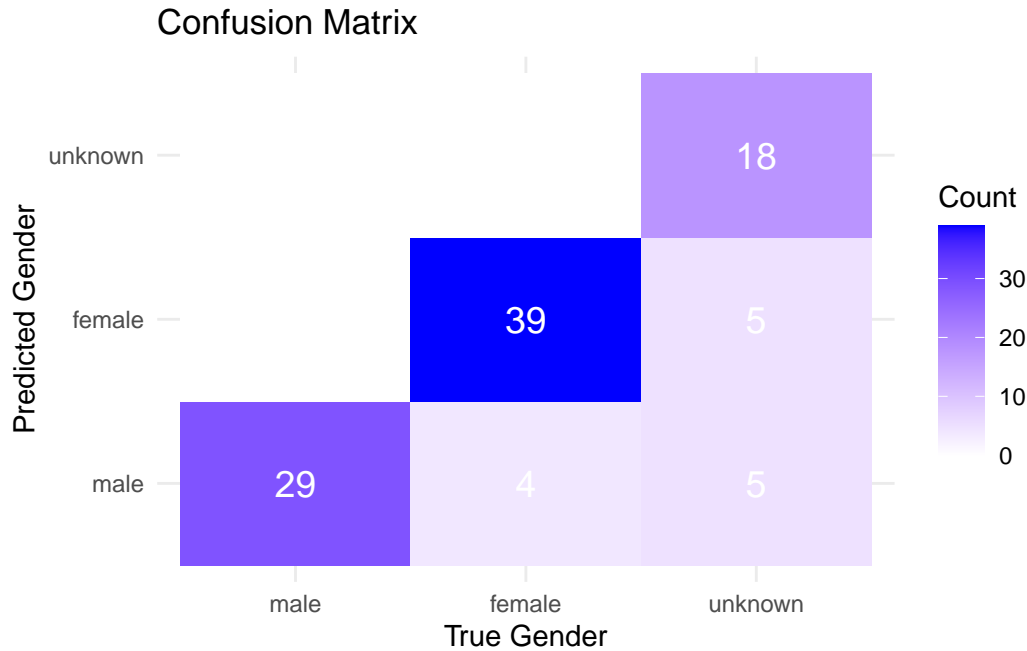
```

#This is the output of the confusion matrix.

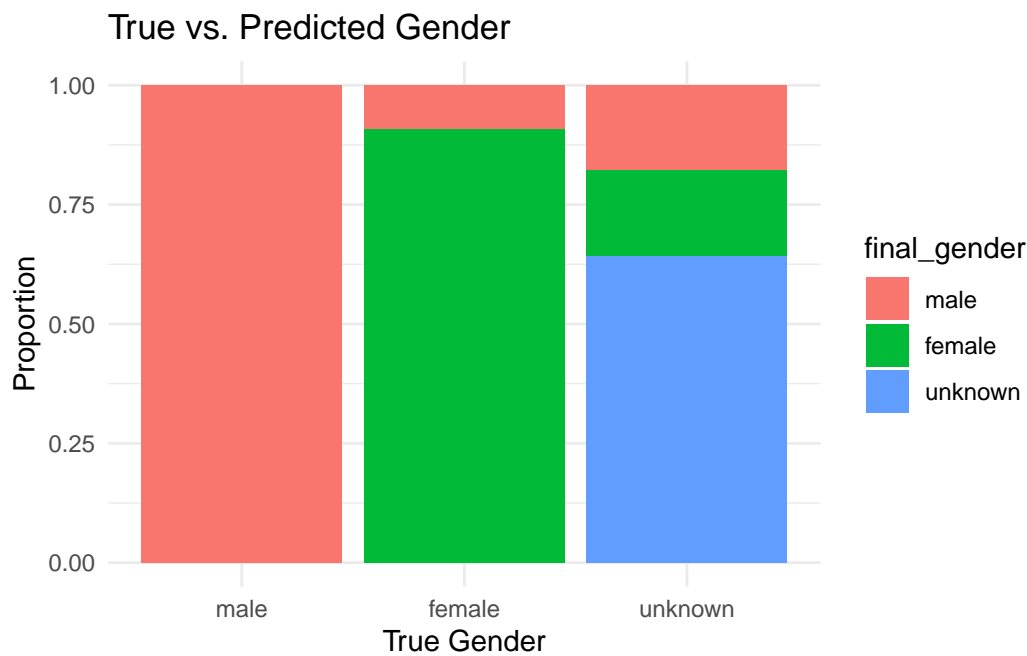
# Convert to data frame for ggplot2
conf_df <- as.data.frame(table(true_gender$final_gender, true_gender$true_gender))
colnames(conf_df) <- c("Predicted", "True", "Count")

# Plot confusion matrix as heatmap
ggplot(conf_df, aes(x = True, y = Predicted, fill = Count)) +
  geom_tile() +
  geom_text(aes(label = Count), color = "white", size = 5) +
  scale_fill_gradient(low = "white", high = "blue") +
  labs(title = "Confusion Matrix", x = "True Gender", y = "Predicted Gender") +
  theme_minimal()

```



```
ggplot(true_gender, aes(x = true_gender, fill = final_gender)) +
  geom_bar(position = "fill") + # Proportion of each category
  labs(title = "True vs. Predicted Gender", x = "True Gender", y = "Proportion") +
  theme_minimal()
```



The plot displays the proportional relationship between true and predicted gender labels in the test dataset. For both the “male” and “female” true gender categories, the model performs strongly, with

most predictions aligning correctly with the true labels (e.g., nearly 100% of true “male” cases are predicted as “male”, and most “female” cases are predicted as “female”).

However, for the “unknown” true gender category, the model shows more variability. While the majority of “unknown” cases are correctly predicted as “unknown,” a portion is misclassified as either “male” or “female.” This confirms earlier concerns that the model tends to prefer assigning a binary gender when uncertain, rather than consistently labeling these cases as “unknown.”

Overall, this plot highlights the model’s good performance on clear-cut cases (“male” and “female”) but also its tendency to underperform on #ambiguous or hard-to-classify usernames, represented by the “unknown” group.

```
# It is highly recommended to avoid running this section, as it is
#extremely computationally expensive.

#Apply the procedure to the final dataset.
f_1 <- final_dataset[1:5000,]
df_2 <- final_dataset[5001:10000, ]
df_3 <- final_dataset[10001:15000, ]
df_4 <- final_dataset[15001:20000, ]
df_5 <- final_dataset[20001:25000, ]
df_6 <- final_dataset[25001:30000, ]
df_7 <- final_dataset[30001:35000, ]
df_8 <- final_dataset[35001:40000, ]
df_9 <- final_dataset[40001:nrow(final_dataset), ]

df_1_gender <- apply_advanced_gender_detection(df_1, "authorDisplayName")
df_1_final <- process_unknown_names(df_1_gender, batch_size = 50)
write.xlsx(df_1_final, file = "C:/Users/schia/Documents/LMU/CSS/df_1_final.xlsx")

df_2_gender <- apply_advanced_gender_detection(df_2, "authorDisplayName")
df_2_final <- process_unknown_names(df_2_gender, batch_size = 50)
write.xlsx(df_2_final, file = "C:/Users/schia/Documents/LMU/CSS/df_2_final.xlsx")

df_3_gender <- apply_advanced_gender_detection(df_3, "authorDisplayName")
df_3_final <- process_unknown_names(df_3_gender, batch_size = 50)
write.xlsx(df_3_final, file = "C:/Users/schia/Documents/LMU/CSS/df_3_final.xlsx")

df_4_gender <- apply_advanced_gender_detection(df_4, "authorDisplayName")
df_4_final <- process_unknown_names(df_4_gender, batch_size = 50)
write.xlsx(df_4_final, file = "C:/Users/schia/Documents/LMU/CSS/df_4_final.xlsx")

df_5_gender <- apply_advanced_gender_detection(df_5, "authorDisplayName")
df_5_final <- process_unknown_names(df_5_gender, batch_size = 50)
write.xlsx(df_5_final, file = "C:/Users/schia/Documents/LMU/CSS/df_5_final.xlsx")
```

```

df_6_gender <- apply_advanced_gender_detection(df_6, "authorDisplayName")
df_6_final <- process_unknown_names(df_6_gender, batch_size = 50)
write.xlsx(df_6_final, file = "C:/Users/schia/Documents/LMU/CSS/df_6_final.xlsx")

df_7_gender <- apply_advanced_gender_detection(df_7, "authorDisplayName")
df_7_final <- process_unknown_names(df_7_gender, batch_size = 50)
write.xlsx(df_7_final, file = "C:/Users/schia/Documents/LMU/CSS/df_7_final.xlsx")

df_8_gender <- apply_advanced_gender_detection(df_8, "authorDisplayName")
df_8_final <- process_unknown_names(df_8_gender, batch_size = 50)
write.xlsx(df_8_final, file = "C:/Users/schia/Documents/LMU/CSS/df_8_final.xlsx")

df_9_gender <- apply_advanced_gender_detection(df_9, "authorDisplayName")
df_9_final <- process_unknown_names(df_9_gender, batch_size = 50)
write.xlsx(df_9_final, file = "C:/Users/schia/Documents/LMU/CSS/df_9_final.xlsx")

```

```

#Emoji translation using the chat gpt API
# usethis::edit_r_environ(scope = "user")
chat_key = Sys.getenv("chat_key")

translate_batch <- function(texts) {
  response <- POST(
    url = "https://api.openai.com/v1/chat/completions",
    add_headers(
      `Authorization` = paste("Bearer", chat_key),
      `Content-Type` = "application/json"
    ),
    body = toJSON(list(
      model = "gpt-4o-mini",
      messages = list(
        list(role = "system", content = "You are an expert in emoji
        descriptions and their translations into Italian.
        I will provide you with a character vector containing English descriptions.
        Your task is to return a character vector of the same length, with each
        description accurately translated into Italian.
        Here are the English descriptions:
        %s
        Please provide ONLY the Italian translations, one per line, with no
        explanations, numbering, prefixes or additional text.
        Do not include the original English text or any other commentary."),
        list(role = "user", content = paste(texts, collapse = "\n"))
      ),
      temperature = 0.3
    ), auto_unbox = TRUE)

```

```

)

# Extract response
result <- content(response, as = "parsed", encoding = "UTF-8")

# Return split translations
if (!is.null(result$choices)) {
  return(strsplit(result$choices[[1]]$message$content, "\n")[[1]])
} else {
  return(rep(NA, length(texts)))
}
}

# Split into chunks of 50 and process in batches
batch_size <- 50
num_batches <- ceiling(length(emoji.descriptions) / batch_size)

#Run batch processing
emoji_translations <- unlist(lapply(seq(1, length(emoji.descriptions),
                                     by = batch_size), function(i) {
  end <- min(i + batch_size - 1, length(emoji.descriptions))
  print(paste("Translating batch", i, "to", end))
  translate_batch(emoji.descriptions[i:end])
})))

#Since there were 6 missing translations, those were identified through human detection
emoji.descriptions[c(737,738,3101,3153,3174,3175)]
missing_transl <- c(
  "donna che aggrota le sopracciglia: tono di pelle scura",
  "donna che aggrota le sopracciglia: tono di pelle scura",
  "bacio: donna, uomo, tono della pelle medio-chiaro, tono della pelle medio-scuro",
  "bacio: uomo, uomo, tono della pelle medio-chiaro, tono della pelle medio-scuro",
  "bacio: uomo, uomo, tono della pelle medio-scuro, tono della pelle scuro",
  "bacio: uomo, uomo, tono della pelle medio-scuro, tono della pelle scuro"
)

result <- character(length(emoji_translations) + length(missing_transl))
indices <- rep(NA, length(result))
missing_transl <- indices[c(737,738,3101,3153,3174,3175)]
emoji_translations <- indices[-c(737,738,3101,3153,3174,3175)]
  [is.na(indices[-c(737,738,3101,3153,3174,3175)])]

result <- indices[!is.na(indices)]
write.table(result, file = "emoji_tradotte.txt")

```

```

#Get the NRC dictionary
nrc_lex <- get_sentiment_dictionary(dictionary = "nrc", language = "italian")

#Get the Italian Sentiment Lexicon
system("git clone https://github.com/opener-project/public-sentiment-lexicons.git")
#Clone the repository from the github page

#The files will be in the propagation_lexicons/it folder
lmf_file <- "public-sentiment-lexicons/propagation_lexicons/it/
            it.lemma.sy.an.hypo.rels.maxdepth5.seed500.maj.gold.lmf"
xml_data <- read_xml(lmf_file)

#Using the .lmf extension, first load the words
entry_nodes <- xml_find_all(xml_data, "//LexicalEntry")

#Then load the sentiment
#Select all 'sense' nodes within 'entry' nodes
sense_nodes <- xml_find_all(xml_data, ".//Sense//Sentiment")

# Function to extract lexicon from an LMF file
extract_lexicon <- function(file_path) {
  xml_data <- read_xml(file_path)

  # Extract lexical entries
  entries <- xml_find_all(xml_data, "//LexicalEntry")

  # Create a data frame of words
  words <- lapply(entries, function(entry) {
    # Extract the word form
    word <- xml_text(xml_find_first(entry, ".//Lemma/@writtenForm"))

    sentiment_node <- xml_find_first(entry, ".//Sense/Sentiment")
    sentiment_value <- if(!is.na(sentiment_node)) {
      xml_attr(sentiment_node, "polarity")
    } else {
      # If no explicit value, use the file name to determine sentiment
      str_extract(file_path, "[a-z]+(?:\\.lmf$)")
    }

    return(data.frame(word = word, sentiment = sentiment_value))
  })

  # Combine all words
  result <- do.call(rbind, words)

```

```

    return(result)
}
cnr_lex <- extract_lexicon(lmf_file)

```

```

#|fig-cap: "Most frequent words"
ggplot(highest_freq_words_erased[1:15, ],
       aes(x = reorder(word, -freq), y = freq)) +
  geom_bar(stat = "identity", fill = "lightblue") +
  labs(y = "Word frequencies",
       x = "")
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

