

UNIVERSITÀ DEGLI STUDI DI TRENTO
DIPARTIMENTO di SOCIOLOGIA E RICERCA SOCIALE
SOCIOLOGY AND SOCIAL RESEARCH



Cultural Analytics

Prof. Acerbi Alberto

aa. 2023/2024

Trust in global news
Sentiment analysis of BBC articles

Research report

Giulia Bessegato – giulia.bessegato@studenti.unitn.it

Word count sections (3000 max.): 2654

SUMMARY

1. INTRODUCTION.....	3
2. METHODS.....	3
2.1 The dataset.....	3
2.2 Topic Modeling and Term Frequency-Inverse Document Frequency (TF-IDF)....	4
3. RESULTS.....	5
3.1.1 AFINN lexicon.....	6
3.1.2 NRC Sentiment and Emotion Lexicons.....	7
4. DISCUSSION.....	10
4.1 Limitations.....	11
5. REFERENCES.....	12
6. APPENDIX.....	14

1. INTRODUCTION

The influence of media on society has been investigated and still is. The analysis of media concerns images or videos along with texts, such as tweets, comments, reviews, or articles.

In this context, the analysis concerns journal articles; an important consideration regards the fact that the audience is not reading an objective report of facts, but rather a framed story that includes certain words instead of others.

“A media frame is a written, spoken, graphical, or visual message modality that a communicator uses to contextualize a topic, such as a person, event, episode, or issue, within a text transmitted to receivers by means of mediation” (D’Angelo, 2017, p.1)

This conception of articles' texts on media enhances the importance of the analysis, whereas framing issues and events may configure public opinion (De Vree, 2005) as well as the emotional responses of the reader (de los Santos & Nabi, 2019). The role of emotions in news framing could be considered part of the developments and trends in framing research: indeed, there is a growing corpus of research on media and framing processes influence where emotions play a more central role compared to the past (Schuck & Feinholdt, 2015). This change of focus may be due to an equal attention towards rationality and emotions. Nabi(1999) wrote regarding a complementary approach of persuasion that could sum both the “emotional” and “rational” ones; indeed, the author considers the actor in the process of persuasion as capable of reasoning through emotions. Framing research on news based on emotions emerges particularly in political communication research. In this field, political opinions and behavior are studied predominantly in terms of negative emotions (specifically fear and anger) compared to positive ones (Schuck & Feinholdt, 2015).

In this context, sentiment analysis will be performed on news articles of a global news corporation, the BBC, making room for both negative and positive emotions. In addition to this, other complementary aspects of the texts will be considered, such as the topics of the articles and a deeper analysis of different emotions beyond the negative and positive labels.

2. METHODS

2.1 The dataset

The dataset analyzed includes 2225 BBC news articles, published in 2004 and 2005. The dataset was elaborated by Greene and Cunningham (2006) who extracted the dataset from the BBC’s official website. In order to better conduct the analysis the authors decided to label

each article under one of five categories, that are business, entertainment, politics, sport, and tech.

Having regard to the influence of news on society, an important thing to consider is that this clout depends in part on the economic support and influence that a news brand has (Cissel, 2012). The British Broadcasting Corporation (BBC) has a weekly audience of 318 million people: however, the influence of the brand doesn't stop on the numbers of readers, but instead on the fact that these audiences come from different parts of the world. The corporation is one of the most important news networks that spread global news beyond national service (BBC, 2024): the online BBC World Service is translated into more than forty languages across fifty-nine countries. The article of Dencik (2013) highlights the influence and importance of the affirmation and spread of a specific journalistic and political culture that represents a certain global frame and story.

In the following section, I write more details about the contents of the articles, considering the labeling in the five categories in order to better organize the analysis.

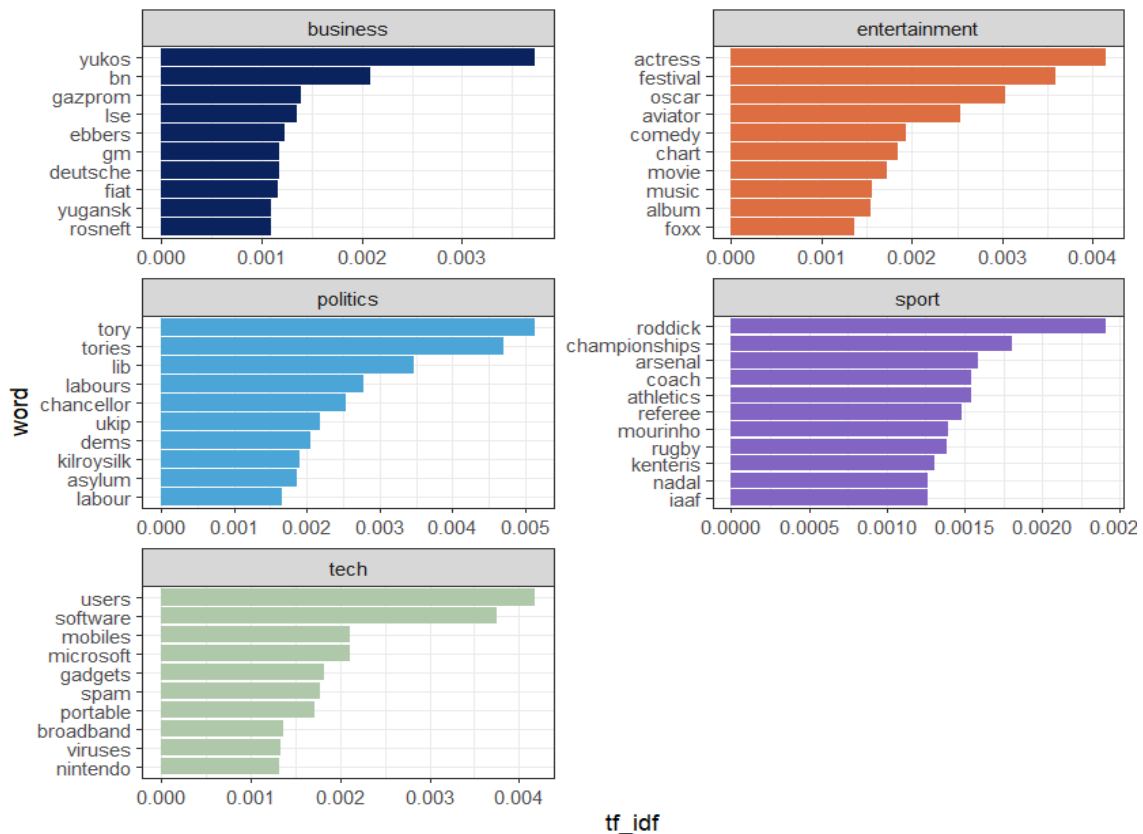
2.2 Topic Modeling and Term Frequency-Inverse Document Frequency (TF-IDF)

I decided to conduct a topic modeling analysis to confirm the categorization already performed in the dataset, by analyzing the contents for each article through a Latent Dirichlet Allocation (LDA) model with a beta matrix that identifies ten main terms that better describe each topic. These terms are graphically represented by graph n.1 in the Appendix section, where the words confirm the categorization performed in Greene and Cunningham's article (2006). After the topic modeling analysis selected ten words presented for each category, I manually labeled the category for each graph.

Although the topic modeling previously presented already identifies common themes in each article category, further analysis has been carried out in order to define the topics most frequently mentioned in the drafting of articles. The analysis regards the Term Frequency - Inverse Document Frequency (TF-IDF), which permits the identification of the most significant words in a document compared to the rest of the corpus, thus, without the influence of prominent words in all the articles. This technique better represents which events and people each article focuses on more, and is graphically reported in graph n.2.

Within the business category, the articles mainly report events related to large oil companies and multinational corporations, while the entertainment words such as actress, festival, and oscar indicate the greater focus on awards and people in the film industry compared to the music one.

Graph n. 2, Score of tf-idf for single words in each category



In the politics category, in 2004 and 2005 the focus is on British political parties, highlighting issues relating to the domestic policy of the UK. Sports articles give attention to global sporting events and prominent personalities, particularly tennis, football, and athletics. Finally, within the tech category, the predominant focus is on the evolution of consumer technology, particularly with an emphasis on mobile devices, software, and cybersecurity.

In the following results section, a sentiment analysis will be conducted, considering different lexicons.

3. RESULTS

3.1 Sentiment analysis

Here below is reported a general sentiment analysis definition by Mohammad(2021).

“Sentiment analysis is an umbrella term for the determination of valence, emotions, and other actual states from text or speech automatically using computer algorithms“ - (Mohammad, 2021, p.1)

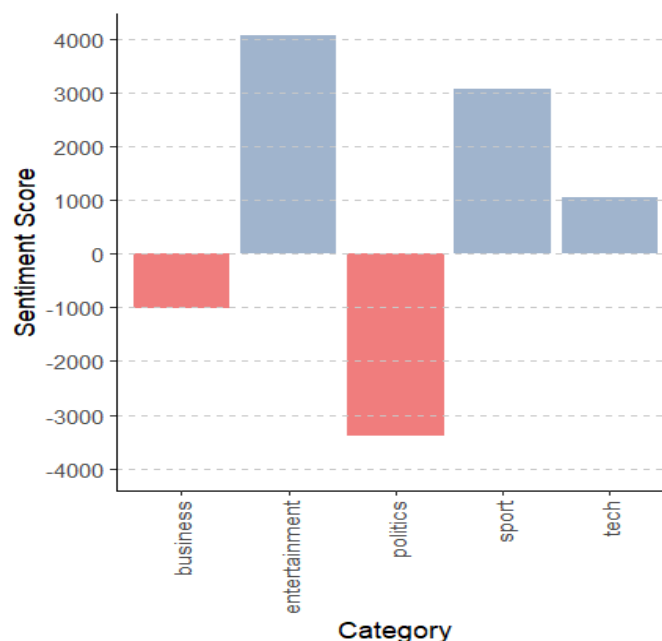
Over the course of the last decade, the research in sentiment analysis has inquired about when a text has a positive or negative polarity (Mohammad &Turney, 2016), indeed, more than one approach has been developed. In this context what will be considered is the Lexicon

Based approach, which uses a list of manually labeled words in order to allocate textual content (Azam & Mehmood, 2020); since many lexicons could be used, I will select two of these.

3.1.1 AFINN lexicon

The first analysis to be carried out involves the AFINN lexicon, which categorizes words as positive or negative, providing a scale in order to differentiate and measure different tones of positivity and negativity of words. This analysis considers the division of the dataset into categories already reported and is graphically disclosed in graph n. 3.

Graph n.3, AFINN lexicon sentiment scores for each category



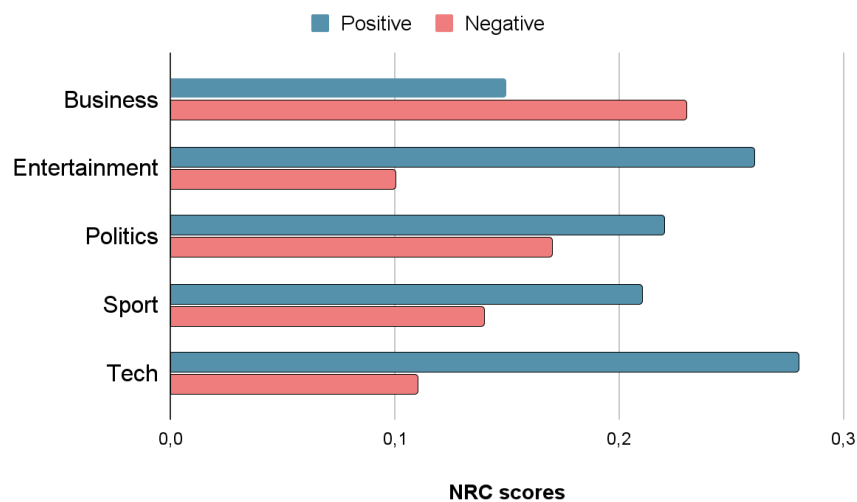
Looking at the total scores in all the datasets, an interesting issue is the fact that while the highest rating of positivity is over 4000 points, the highest rating of negativity is around 3000 points. The categories that present positive scores are entertainment, sport, and tech, while more negative scores are measured within business and politics. The two opposite poles are entertainment for positive scores and politics for negative ones, suggesting that while entertainment articles have a strongly positive tone which may reflect content that elicits pleasant emotions, politics articles tend to deal with more controversial, or potentially distressing topics, resulting from a strongly negative sentiment score compared to the rest of the corpus. Only by looking at these results, the categorization of the articles has been found relevant in measuring sentimental scores of words in the articles.

3.1.2 NRC Sentiment and Emotion Lexicons

Another lexicon that extracts positive and negative emotions from the text is the NRC emotion lexicon. The decision to include both lexicons in the analysis involves two issues. The first is the complementary measure of positive and negative emotions: the comparison of the two can offer a complementary view of the sentiment analysis performed here, indeed, while AFINN is able to better highlight the differences in intensity within the feelings expressed, NRC suggests a more balanced distribution providing a binary overview of sentiment. The second reason regards the fact that the NRC lexicon measures, in addition, eight basic emotions based on the Plutchik plus polarity model, which are anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, linked to 14182 English words (Tabak & Evrim, 2016).

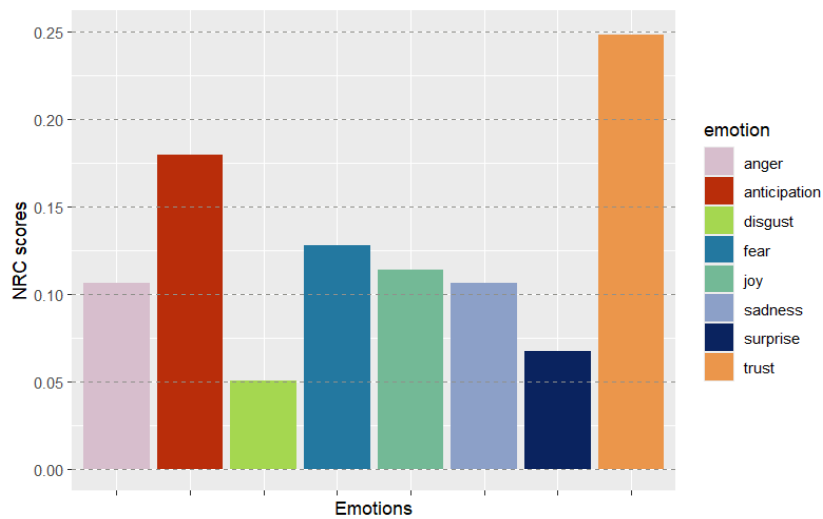
Graph n. 4 shows the scores of negative and positive emotions using the NRC lexicon. The categories with the most negative emotions are business, where the negative score is over the positive, and politics, which has a more balanced score for both emotions.

Graph n.4, NRC emotion lexicon for positive and negative emotions by category



It can be noticed that while the categories that are found to have a higher presence of negative emotions in the AFINN lexicon analysis correspond to the same encountered in the NRC lexicon, in this case, the positive emotion scores prevail in the tech category more than in the entertainment one. This could be explained by the fact that negative emotions are present in higher quantities in the tech category. Graph n. 5 shows the general scores of NRC considering all the BBC articles in the dataset. Since the analysis of positive and negative words has been already performed, these are here excluded.

Graph n. 5, NRC scores of eight emotions



Looking at the predominant emotions compared to the others, trust has been detected to a greater extent indicating a general sense of reliability and safety in the analyzed articles. The second one is anticipation, suggesting a stimulation of reader's expectations and interests. Fear and Joy have very close scores, reflecting a balance between worries and happiness in the contents. Sadness, Anger, and Surprise are less common but still present, and, finally, disgust is the less frequent emotion, suggesting that BBC content may minimize or avoid repulsive reactions.

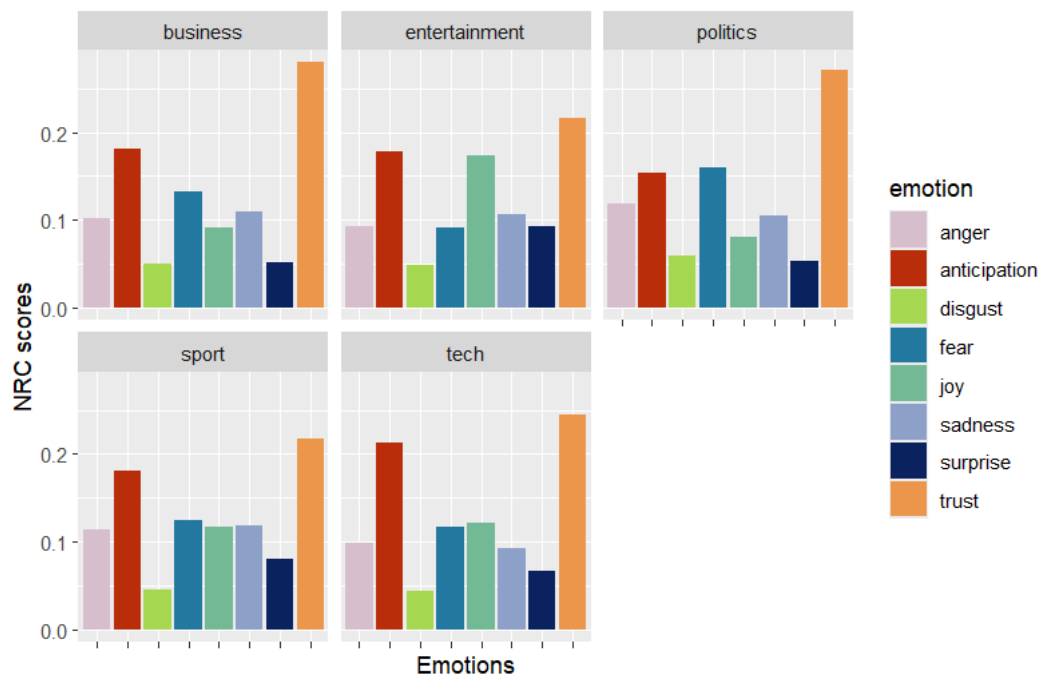
Other considerations regard how these predominant emotions are distributed. Graph n.6 shows the distribution of the emotions taking into consideration the topics of the articles.

The emotion of trust which is the most present in the dataset has the highest scores in the business and politics category, while anticipation is the second most present emotion that is frequent in tech. From both the brand and general level, the emotion of trust has an impact on the role of news inside society. A challenge for news brands in terms of gaining the trust of the audience is to consider criteria that define the "right" and "trusted" information. Examples of these criteria are the sharing of political values, as well as certain characteristics of the public such as gender, income, age, or education.

The emotion of trust which is the most present in the dataset has the highest scores in the business and politics category, while anticipation is the second most present emotion that is frequent in tech. From both the brand and general level, the emotion of trust has an impact on the role of news inside society. A challenge for news brands in terms of gaining the trust of the audience is to consider criteria that define the "right" and "trusted" information.

Examples of these criteria are the sharing of political values, as well as certain characteristics of the public such as gender, income, age, or education.

Graph n. 6, NRC scores for eight emotions per category



Once considered the most prevalent emotions in the articles, the focus can be shifted to what kind of emotions are conferred in each category.

Regarding the business category, the prevalence of anticipation and trust suggests that these articles tend to convey a sense of reliability in writing about companies that are global economic pillars. The creation of confidence in the reader can be found in the politics category as well; however, in this case, the mix of trust, anticipation, and fear may reflect the often uncertain and polarizing nature of political news; confidence may be related to the writing of institutional roles of politics, while the emotion of fear to uncertainties or controversies related to certain politicians and parties.

In the entertainment category that particularly focuses on cultural events of great impact such as festivals or the Oscars night, the component of joy is a major presence compared to other categories, along with a more equilibrium between trust and anticipation.

In the sport category, anticipation is the prevailing emotion. The focus on personalities in the world of sport as well as sporting events indicates a focus on competitions, generating anticipation for results and confidence in the abilities of teams or athletes.

The technology category is dominated by trust and anticipation, suggesting an expectation for new developments or products, in conjunction with a strong sense of reliability in technological information.

4. DISCUSSION

According to the analysis, two main issues linked to each other should be featured. The first is the fact that in the BBC articles, positive emotions prevail over negative ones, in the analysis of both AFINN and NRC lexicons. This positivity is due in part to the fact that the prevalent emotions in the dataset are the ones of trust and anticipation, indeed, the NRC Emotion Lexicon labels them as positive emotions (Mohammad & Turney, 2010; Mohammad & Turney, 2013). This outcome is not surprising: the dataset includes articles from one of the most important worldwide global news brands, that, as previously reported, not only has a national reputation, but also the trust of readers coming from different national backgrounds. In the contemporary era, the availability of news sources has increased, as well as the brand of news. In the report of Northwestern University (Abernathy, 2023) it is expected that by the time 2024 ends, the US will have lost a third of its newspapers since the year of the publication date of the articles considered, that is 2005.

A strong criteria of trust in reading BBC news involve the transparency of information sources; on the official site of BBC, there is a description of the sub-brand BBC Verify, launched in 2023.

“the BBC's news teams are using ever more sophisticated tools, techniques and technology to check and verify videos like the Kremlin drone footage, as well as images and information. They do this to ensure our journalism meets the rigorous editorial standards the BBC is proud to uphold” - (BBC, 2023, Explaining the 'how'—The launch of BBC Verify, <https://www.bbc.com/news/uk-65650822>)

However, the process of selection and avoidance of news brands considered by researchers do not always coincide with the criteria of reliability of the truthfulness of a source (de los Santos & Nabi, 2019) (Newman et al., 2024). In the relationship between journalism and the news media, the public perception of trust is fundamental; indeed, a common viewpoint is a subsequent part of reality. For instance, part of the literature confirms that negative news contents have a more enduring and trusted impact than positive ones (Soroka & McAdams, 2015) (Lindgren et al, 2022). Regarding news about politics, the article by Kleinnijenhuis and

colleagues (2006) reported that the consumption of negative news could negatively affect the trust of the audience in politicians and political parties. Somewhat, the framing of articles evokes emotions in the reader, who in turn influences his/her decision to continue reading.

The second relevant issue regards the division of the articles into categories that better represent the distribution of emotions in the dataset. The prevalent emotion of trust has the highest scores in the business and politics categories, nevertheless the highest negativity scores within their articles.

I decided to focus on the categories of politics and business as long as in the dataset the ambivalence of emotions is stronger, indicating that companies, singular personalities, and political parties, are negatively framed while named over and over, compared to other news topics. This final discussion section pays attention to the tendency of framing business and politics articles combining levels of trust and negativity. In making both these emotions prevalent, the BBC guarantees itself a bond with the audience, using the criteria of both the reliability of news sources and the presence of negative emotions.

4.1 Limitations

Regarding the limitations of this work, three issues should be cited. First of all, the dataset includes articles not published in recent years, and the fact that the pandemic crisis has changed the news context could not be considered by this analysis. The second is the fact that there is no specific information on the articles selected in the dataset, considering the fact that BBC publishes a huge number of articles every year and that there are many subgroups of the company. Finally, this analysis is only text-based, while sentiment and framing analysis could include images and video of news as well.

5. REFERENCES

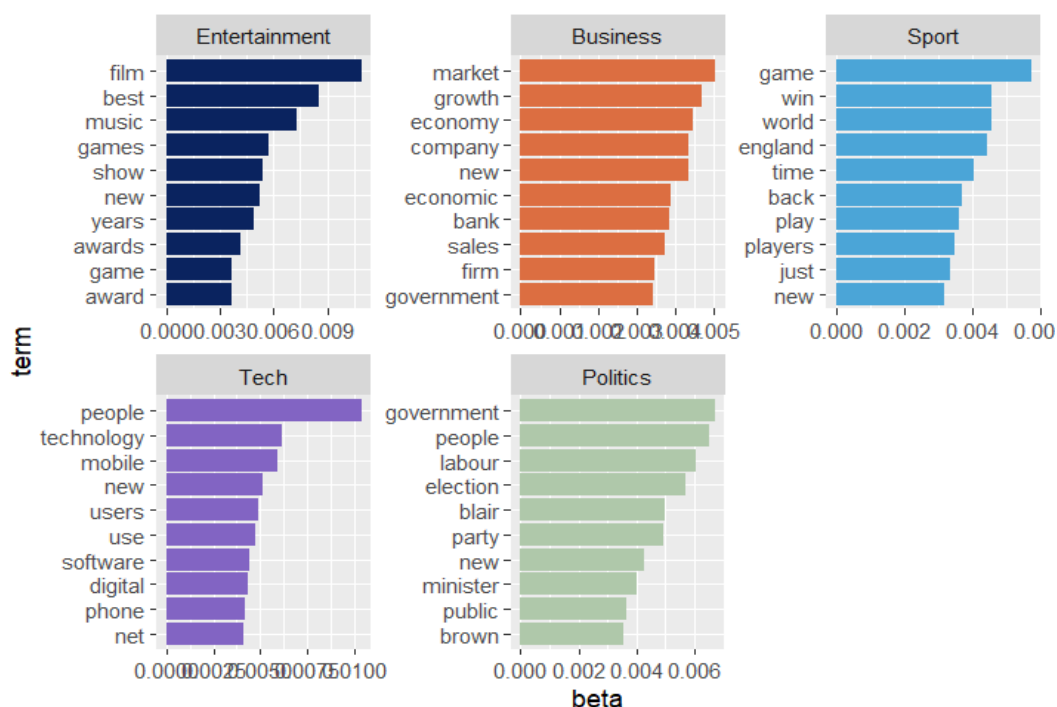
- Abernathy, P. M. (2023). The state of local news: The 2023 report. Medill Local News Initiative. Northwestern University.
<https://localnewsinitiative.northwestern.edu/projects/state-of-local-news/2023/report/>
- Azam, N., Tahir, B., & Mehmood, M. A. (2020). Sentiment and emotion analysis of text: A survey on approaches and resources. *Language & Technology*, 87.
- Cissel, M. (2012). Media Framing: a comparative content analysis on mainstream and alternative news coverage of Occupy Wall Street. *The Elon Journal of Undergraduate Research in Communications*, 3(1), 67-77.
- D'Angelo, P. (2017). Framing: media frames. *The international encyclopedia of media effects*, 2, 634-644.
- de los Santos, T. M., & Nabi, R. L. (2019). Emotionally Charged: Exploring the Role of Emotion in Online News Information Seeking and Processing. *Journal of Broadcasting and Electronic Media*, 63(1), 39–58.
- De Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal+ Document Design*, 13(1), 51–62.
- Dencik, L. (2013). What global citizens and whose global moral order? Defining the global at BBC World News. *Global Media and Communication*.
<https://doi.org/10.1177/1742766513479716>
- BBC (2023) *Explaining the 'how'—The launch of BBC Verify*.
<https://www.bbc.com/news/uk-65650822>
- BBC (2024) *Global news services*. (n.d.). Retrieved 29 August 2024, from
<https://www.bbc.co.uk/aboutthebbc/whatwedo/worldservice/>
- Greene, D., & Cunningham, P. (2006). Practical solutions to the problem of diagonal dominance in kernel document clustering. *Proceedings of the 23rd International Conference on Machine Learning*, 377–384. <https://doi.org/10.1145/1143844.1143892>
- Kleinnijenhuis, J., Van Hoof, A. M., & Oegema, D. (2006). Negative news and the sleeper effect of distrust. *Harvard International Journal of Press/Politics*, 11(2), 86-104.
- Lindgren, E., Lindholm, T., Vliegthart, R., Boomgaarden, H. G., Damstra, A., Strömbäck, J., & Tsifti, Y. (2022). Trusting the facts: The role of framing, news media as a (trusted) source, and opinion resonance for perceived truth in statistical statements. *Journalism & Mass Communication Quarterly*, 10776990221117117.
- Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.
<https://doi.org/10.1111/j.1467-6427.2012.00460.x>
- Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL-HLT*

- 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text (pp. 26-34). <https://www.aclweb.org/anthology/W10-2004>
- Mohammad, S. M. (2021). *Sentiment Analysis: Automatically Detecting Valence, Emotions, and Other Affectual States from Text* (No. arXiv:2005.11882). arXiv. <https://doi.org/10.48550/arXiv.2005.11882>
- Nabi, R. L. (1999). A Cognitive-Functional Model for the Effects of Discrete Negative Emotions on Information Processing, Attitude Change, and Recall. *Communication Theory*, 9(3), 292–320. <https://doi.org/10.1111/j.1468-2885.1999.tb00172.x>
- Newman, N., Fletcher, R., Robertson, C. T., Arguedas, A. R., & Nielsen, R. K. (2024). *Reuters Institute Digital News Report 2024*.
- Priyavrat, & Sharma, N. (2018). Sentiment Analysis using tidytext package in R. *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 577–580. <https://doi.org/10.1109/ICSCCC.2018.8703296>
- Tabak & Evrim (2016) Comparison of emotion lexicons. In Proceedings of the 2016 HONET-ICT, Nicosia, Cyprus, pp. 154–158.
- Soroka, S., & McAdams, S. (2015). News, politics, and negativity. *Political communication*, 32(1), 1-22.
- Schuck, A., & Feinholdt, A. (2015). *News Framing Effects and Emotions* (pp. 1–15). <https://doi.org/10.1002/9781118900772.etrds0241>
- Skipworth, S. (2011). *The Consequences of Trust in the Media: A Moderating Effect of Media Trust on the Relationship between Attention to Television and Political Knowledge* (SSRN Scholarly Paper No. 1766887). <https://papers.ssrn.com/abstract=1766887>

6. APPENDIX

6.1 Charts

Graph n.1, Topic modeling analysis on the dataset



6.2 Attached file

- *bbc_cvs_converted*

The original dataset has been converted in a csv file with columns ordered and labeled for each article. The folder contains the original dataset and the code used to convert the dataset used in the R file.

- *Bessegato_Giulia_CulturalAnalysis.R**

RStudio file