# An Automated Sentiment Analysis of The Sun and USA Today's Anti-Trump Protests News Articles in 2018

**Dr. Sara Mohamed Awad**
*Assistant Professor*
*Department of Languages,*
*Sadat Academy for Management Sciences, Egypt*
*awad.sara80@gmail.com*

## Abstract

Tracing sentiments in discourse has been grabbing the attention of both researchers and professionals and since it can provide clues to views, attitudes, prediction of reactions, etc. The use of sentiments in news coverage texts is to be attended to since it, definitely, has an impact on the public opinion. Adopting sentiment analysis, this study examines the use of sentiments in the news articles of the two most circulated UK and U.S. newspapers during 2018-2019, namely The Sun and USA Today [online versions], as delivering the same news of the anti-Trump protests in the UK and the USA associated with the president's visit to the UK in 2018. The analysis, following the lexicon-based approach, is conducted in three stages: 1. subjectivity detection; 2. polarity (i.e. semantic orientation) identification; 3. magnitude (i.e. intensity) assessment. Based on the analysis conducted, a conclusion is made that both newspapers articles on the issue are generally objective and that USA Today's use of subjective language is more restricted in volume yet more sentimental.

**Keywords:**
sentiments, sentiment analysis, subjectivity, polarity, magnitude.

# تحليل آلي للمشاعر المتضمنة بالتغطية المقالية الإخبارية لجريدتي "ذا صن" و"يو إس إيه توداي" لمظاهرات معارضة لترامب في 2018م

## د. سارة محمد عوض

**مدرس بقسم اللغات- كلية العلوم الإدارية- أكاديمية السادات للعلوم الإدارية- مصر**

**الملخص العربي:**

أضحت دراسة المشاعر بالخطاب مسيطرة على انتباه كل من الباحثين والمهنيين, حيث أنه بإمكانها الكشف عن آراء واتجاهات وكذلك المساعدة في تنبوء ردود أفعال مستقبل الخطاب وغير ذلك. ولا يستثنى هنا بل ينبغي الانتباه إلى استخدام المشاعر في نصوص التغطية الإخبارية نظراً لأهميتها في التأثير على الرأي العام. وعليه تقوم هذه الدراسة بتوظيف التحليل الآلي للمشاعر لتتبع استخدام المشاعر في المقالات الإخبارية لاثنين من أكثر الصحف تداولاً خلال الفترة من 2018 حتى 2019 في المملكة المتحدة والولايات المتحدة الأمريكية ، ألا وهما "ذا صن" و"يو إس إيه توداي" (النسخة الالكترونية)، في تغطيتهما لأنباء المظاهرات المعارضة لترامب في المملكة المتحدة وكذلك الولايات المتحدة والمرتبطة بزيارة الرئيس للمملكة المتحدة في 2018. هذا ويجرى التحليل, معتمداً على النمط المعجمي، على ثلاث مراحل: أولاً, تتبع ذاتية المقالات؛ ثانياً, تحديد الاتجاه الدلالي القطبي؛ ثالثاً, قياس القوة الدلالية. واستناداً إلى نتائج التحليل فقد استخلص البحث أن استخدام المشاعر بالمقالات الإخبارية موضع الدراسة محدود وتزداد محدوديته في التغطية الأمريكية إلا أنها، على الرغم من ذلك، تمثل الوزن الدلالي الأثقل.

**كلمات مفتاحية:**

المشاعر، تحليل المشاعر، الذاتية، التوجه الدلالي القطبي، القوة الدلالية.

## Introduction

Tracing sentiments in language literature and use has been receiving much attention in both academia and industry. They are, by all means, key clues to views, attitudes, social relations, etc. Therefore, sentiments may be studied for their current indications or implications, potential efficiency at serving prediction purposes and potential capacity to act as a tool for initiating, boosting or defying a notion, attitude, etc. What is noteworthy is that sentiments may vary from one culture to another (i.e. across societies or even across communities within societies) in terms of their use, frequency of use, polarity type (i.e. positive vs. negative) and intensity. The use of sentiments in news delivery texts is to be attended to since it, definitely, has an impact on the public opinion. Detecting sentiment use in such texts is significant for identifying, tracing and predicting the public opinion and the mood trends of the news text readers.

## Objective

Within a computational linguistics framework, this study examines the use of sentiments in the news articles of the-ranking-first-by-circulation UK and U.S. newspapers during 2018-2019, namely *The Sun* and *USA Today* [online versions], according to Statista.com, Agilitypr.com and Worldatlas.com, as delivering the same news of the anti-Trump protests in the UK and the USA associated with the president's visit to the UK in 2018.

## Research questions

The study answers the following questions:
1. Are the protests news articles of the two newspapers generally subjective or objective?
2. Is the overall semantic orientation (i.e. polarity) of the articles positive or negative?
3. How strong are the sentiments used in the articles of the two newspapers?

**Previous studies**

Most studies on sentiments in news articles are centered on business-related issues, such as the impact of news on stock prices (Engle, 2004; Devitt & Ahmad, 2007; Tetlock, 2007; Li, Xie, Chen, Wang & Deng, 2014; Remus, Ahmad & Heyer, 2009). Studying sentiments in news coverage still needs more academic attention. Magadza, Mukwazvure, Supreethi, (2014) argue that this is perhaps due to the common belief that journalists tend to be objective. Research conducted in this area includes De Fortuny, De Smedt, Martens & Daelemans (2012), Magadza, Mukwazvure & Supreethi, (2014) and Hoffman (2015) and Lindell & Sartoretto (2018).

**Data & Methodology**

26 articles tackling the issue of the anti-Trump protests associated with the President's visit to the UK in 2018 are collected from the online versions of USA Today and The Sun and are computationally analysed by means of sentiment analysis. Sentiment analysis, according to Brunova (2013), is "one of the rapidly developing methods of natural language processing." Liu (2010) argues that it is the computational study of opinions, sentiments and emotions expressed in text." The analysis broadly classifies textual information into two main categories: objective and subjective. It may go further and detect semantic orientation (i.e. polarity) classifying the subjective data as expressing positive or negative sentiment. Sentiments, however, vary in magnitude (i.e. intensity). They can be strong, medium or weak with a scaled range (Lu, Kong, Quan, Liu & Xu, 2010). Therefore, a final step of such quantitative analysis is to assess magnitude. There are two main approaches to sentiment analysis: a lexicon-based approach and a machine-learning one (Brooke, 2009; Taboada, 2009; Taboada, Brooke, Tofiloski, Voll & Stede, 2011; Banea, Mihalcea & Wiebe, 2011; Moreo, Romeo, Castro & Zurita, 2012; Di Caro & Grella, 2012; Mejova, 2012).

The lexicon-based approach involves constructing a sentiment lexicon as a prior step to conducting analysis (Brooke, 2009), or

using an already existing one. A sentiment lexicon is a list of words and phrases (Banea et al., 2011; liu, 2010) annotated for subjectivity, polarity, or both (Brunova, 2013; Miner, Elder, Hill, Nisbet, Delen & Fast, 2012; Banea et al., 2011; Di Caro & Grella, 2012; Shaikh, Prendinger & Ishizuka, 2008). The simplest lexicons provide binary classification of their entries into "objective" vs. "subjective" or "positive" vs. "negative". Fuzzy lexicons, however, go further and annotate the classified word or phrase with a score conveying the intensity of its polarity (Miner et al., 2012). Machine learning, on the other hand, is an artificial intelligence area. It depends on "induction algorithms and other algorithms that can be said to 'learn'" (Kohavi & Provost, 1998). This approach allows initial sets of seed words to be used for training an artificial intelligence system on processing automated classification of other data sets (Awad, 2017).

Following the lexicon-based approach, the analysis in this study is conducted in three stages: 1. computationally investigating how subjective both the British and American newspaper articles under analysis are; 2. determining the semantic polarity of the article sentiments (i.e. positive vs. negative); 3. assessing the intensity (magnitude) of the sentiments used. The sentiment lexicon used is the one by Mudinas (2012). The lexicon provides a single list of 7038 positive and negative words whose magnitude is assigned according to a weight scale ranging from -5 (a word with the strongest negative weight) to +5 (a word with the strongest positive weight) (Awad, 2017).

A number of necessary pre-processing procedures are applied to the data collected to guarantee accuracy and objectivity of the analysis.

## Data Pre-processing Procedures

A number of necessary pre-processing procedures have been applied to the articles collected in order to make the data computationally analysable. The procedures are as follows:

1. Removing all the parts irrelevant to the article text, such as advertisements, editing details, etc.
2. Converting the texts into a machine-readable format (MRF), which is necessary for the automated text processing required for the computational analysis of the data.
3. Converting the lexicon used into an MRF.

**Analysis:**

The analysis is conducted in three stages. The first round of the analysis of the protest news (i.e. subjectivity detection) yields the findings illustrated in Figure 1 below.
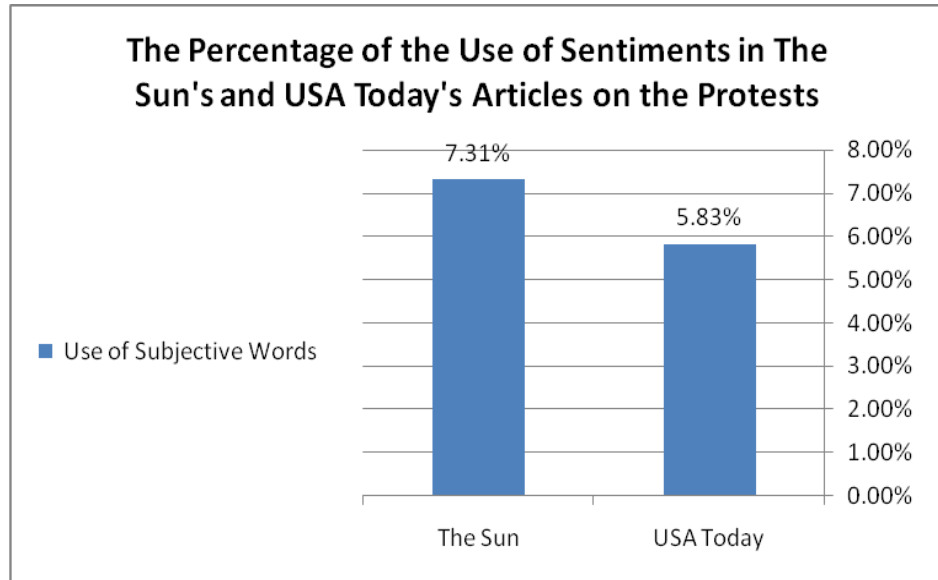


*Figure 1*: Percentage of the use of sentiments in The Sun and USA Today's Articles on the Protest

As shown in Figure 1 above, the use of sentiments is limited in both newspapers (i.e. The Sun: 7.31%; USA Today: 5.83%) being more extensive, however, in The Sun's articles. The second round of the analysis, examining the polarity of the articles based on Mudinas' lexicon (2012), yields the findings in Table 1 below.

Table 1

*Frequency of the Positive and Negative Sentiments in the The Sun and USA Today's Articles Covering the Protests*

| Polarity | The Sun | USA Today |
|---|---|---|
| **Positive Sentiments** | 0.0218 | 0.0339 |
| **Negative Sentiments** | 0.0211 | 0.0291 |

The findings show that positive sentiments are more frequent than negative ones in the articles of the two newspapers. Figure 2 (i.e. 2.a and 2.b) below illustrates the ratio of the positive sentiments to the negative ones in each newspaper respectively.
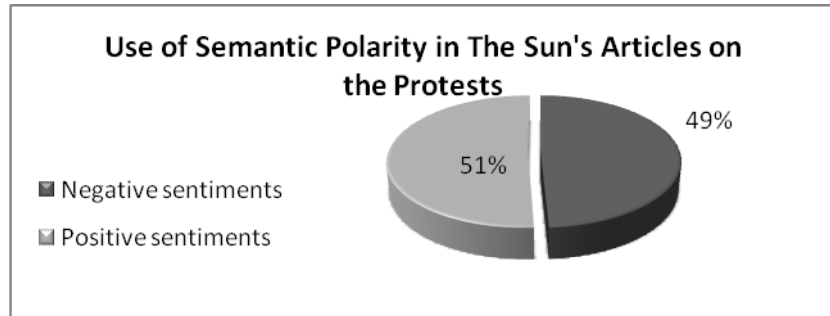


*Figure 2*.a. Use of semantic polarity in The Sun's articles on Protest
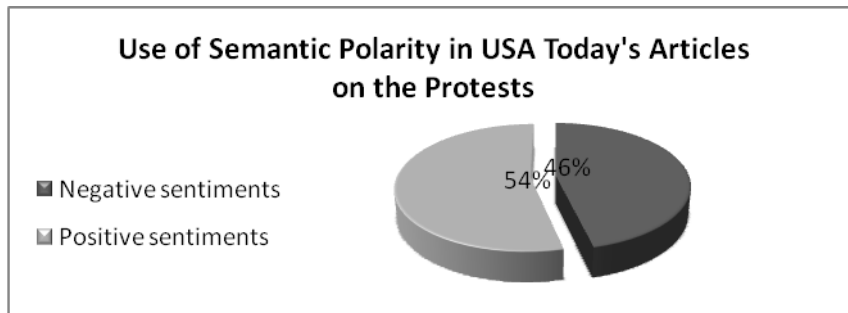


*Figure 2.b. Use of semantic polarity in USA Today's articles on Protest*

Figure 2.a shows that in The Sun's articles on the protests, the percentage of the positive sentiments used is close to that of the negative detected (51% positive; 49% negative) with the positive polarity being 2% greater. Figure 2.b above reveals that the use of positive sentiments (54%) used by USA Today on the topic is much heavier that of the negative ones (46%) with the polarity used being more positive and less negative than in The Sun.

The third round of the analysis involves two procedure: (1) examining the magnitude of the positive and negative sentiments used in the articles by calculating their weights based on Mudinas' (2012) lexicon numerical scale; (2) examining the magnitude of the intensifiers used by calculating their weights based on Brooke's (2009) English intensifiers magnitude numerical scale. Applying the first procedures yields the findings listed in Table 2 below.

Table 2
*Magnitude of the Positive and Negative Sentiments Used in The Sun and USA Today's Articles Covering the Anti-Trump Protest News*

| Mudinas' Numerical Scale | The Sun | | USA Today | |
|---|---|---|---|---|
| | Frequency | Weight | Frequency | Weight |
| -5 | 0 | 0 | 0 | 0 |
| -4 | 12 | -48 | 6 | -24 |
| -3 | 17 | -51 | 46 | -138 |
| -2 | 100 | -200 | 89 | -178 |
| -1 | 43 | -43 | 53 | -53 |
| Total weight of negative sentiments | | -342 | | -393 |
| 1 | 169 | 169 | 217 | 217 |
| 2 | 50 | 100 | 104 | 208 |
| 3 | 15 | 45 | 22 | 66 |
| 4 | 0 | 0 | 2 | 8 |
| 5 | 0 | 0 | 0 | 0 |
| Total weight of positive sentiments | | 314 | | 499 |

Table 2 Cont.

Based on the data in Table 2 above, the following observations are made:

1. The negative sentiments outweigh the positive ones in The Sun's articles while being outweighed by the positive sentiments in USA Today's articles.
2. Both the positive and negative sentiments used in USA Today's articles on the protests outweigh those used in The Sun's, which means that the sentiments used by USA Today are stronger despite the fact that the newspaper's use of subjective language is 1.48% less extensive than the other's.
3. The magnitude of the sentiments used by both newspapers does not reach Mudinas' scale maximum (5) and minimum (-5) values.
4. The positive sentiments used by The Sun are medium strong. They stop at value 3 on the scale. On the other hand the positive sentiments used by USA Today go slightly above the medium level and reach value 4 on the scale, but with the least positive sentiment frequency detected in the articles of both newspapers.
5. The negative sentiments used by both newspapers go slightly above the medium level stopping at value 4 on the scale with low frequencies (i.e. The Sun: 12; USA Today: 6) double higher in The Sun's articles.

Under the second procedure of the third round, intensifiers are classified into amplifiers, which increase the intensity of the sentiments used and are represented by the scale positive values (e.g. "very"), and downtoners which decrease the intensity of the sentiments used and are represented by the scale negative values (e.g. "slightly"). Applying the second procedure (i.e. assessing the magnitude of the intensifiers used in the articles) yields the listed in Table 3 below.

Table 3
*Magnitude of the Intensifiers Used in The Sun and USA Today's Articles Covering the Anti-Trump Protest News*

| Intensifier Type | USA Today Scale Value | Frequency | Weight |
|---|---|---|---|
| **Amplifiers** | 0.25 | 5 | 1.25 |
| | 0.05 | 1 | 0.05 |
| | 0.2 | 1 | 0.2 |
| | 0.3 | 15 | 4.5 |
| | 0.5 | 1 | 0.5 |
| **Total weight of amplifiers** | | | 6.5 |
| **Downtoners** | -0.2 | 7 | -1.4 |
| | -1.5 | 1 | -1.5 |
| Total weight of downtoners | | | -2.9 |

**The Sun**

| | Scale Value | Frequency | Weight |
|---|---|---|---|
| **Amplifiers** | 0.3 | 10 | 3 |
| | 0.2 | 5 | 1 |
| | 0.25 | 3 | 0.75 |
| | 0.5 | 1 | 0.5 |
| **Total weight of amplifiers** | | | 5.25 |
| **downtoners** | -1.5 | 1 | -1.5 |
| | -0.2 | 5 | -1 |
| **Total weight of downtoners** | | | -2.5 |

Based on Table 3, the following observations are made:

1. In the articles of both newspapers, the use of amplifiers outweigh that of downtoners.
2. USA Today's use of amplifiers and downtoners slightly outweigh The Sun's.

## Conclusion

Based on the findings above, it can be stated that The Sun and USA Today, the ranked first by circulation newspapers in the UK and the USA, are generally objective. The analysis of the sentiments used shows that the overall semantic orientation of both newspapers articles on the issue is positive with USA Today being more positive due to two reasons: 1. The ratio of positive sentiment use; 2. The Sun's positive sentiments being outweighed by its negative ones. It also shows that The Sun's sentiments are generally medium strong while USA Today's are more intensive (one degree above medium strong). Thus, it can be concluded that covering the issue of anti-Trump protests associated with the President's visit to the UK in 2018, the American newspaper USA Today's subjective language is more restricted in volume yet more sentimental than The Sun's.

## References

Agility PR Solutions LLC. [Database website] (2019). Top 10 U.K. newspapers by circulation. Retrieved from <https://www.agilitypr.com>. Last accessed on 01/10/2019.

Awad, S. (2013). *Using sentiments in political campaigns: Comparing Obama's 2008 and 2012 presidential election speeches* (Doctoral dissertation). Ain Shams University. Egypt.

Banea, C., Mihalcea, R., & Wiebe, J. (2011). Multilingual sentiment and subjectivity analysis. Retrieved from

<https://people.cs.pitt.edu/~wiebe/pubs/papers/multilingualSubjBookChap2011.pdf>. Last accessed on 01/10/2019.

Brooke, J. (2009). *A Semantic approach to automated text sentiment analysis* (Master's thesis). Simon Fraser University. Canada. ProQuest.

Brunova, E. (2013). Parametrical words in the sentiment lexicon. *International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE), 1* (2)*,* 57-64. Retrieved from <https://www.academia.edu>  . Last accessed on 01/10/2019.

De Fortuny, J., De Smedt, T., Martens, D. & Daelemans, W. (2012). Media coverage in times of political crisis: A text mining approach. *Expert Systems with Applications (ESWA)*, *39*, 11616-11622. ELSAVIER. Retrieved from <https://www.academia.edu>  . Last accessed on 01/10/2019.

Di Caro, L., & Grella, M. (2012). Sentiment analysis via dependency parsing, *Computer Standards & Interface*s.  Retrieved from <https://www.academia.edu>  . Last accessed on 01/10/2019.

Hoffman, A. (2015). Negative and positive sentiments in news coverage of the Boston Marathon bombing. Midwest Political Science Association (MPSA) 2015 Conference. Retrieved from <https;//www.mpsanet.org/Publications-and-Research/MPSA-Conference-Paper-Archive>. Last accessed on 01/10/2019.

Kohavi, R., & Provost, F. (1998). Machine learning - Special issues on applications of machine learning and knowledge discovery process: Glossary of terms, *Machine Learning Journal*, *30* (2-3), 271-274. Boston, MA: Kluwer Academic Publishers. Retrieved from <http://www.robotics.stanford.edu>.
Last accessed on 01/10/2019.

Li, X., Xie, H., Chen, L., Wang, J. & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14-23. ELSAVIER. Retrieved from <https://www.academia.edu>  . Last accessed on 01/10/2019.

Lindell, J. & Sartoretto, P.  (2018). Young people, class and the news: Distinction, socialization and moral sentiments. *Journalism Studies*, *19*, (14). 2042-2061. Taylor & Francis Online. Retrieved

مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية (مجلة علمية محكمة)

from Retrieved from <https://www.tandfonline.com> . Last accessed on 01/10/2019.

Liu, B. (2010). Sentiment analysis and subjectivity. In Indurkhya, N. and Damerau, F. (Eds.). *Handbook of Natural Language Processing* (2nd ed.), 1-38. Retrieved from <https://www.academia.edu> . Last accessed on 01/10/2019.

Lu, Y., Kong, X., Quan, X., Liu, W., & Xu, Y. (2010). Exploring the sentiment strength of user reviews. In Chen, L., Tang, Ch., Yang, J., Gao, Y. (Eds.). *Web-Age: Information Management: Proceedings of the 11th international conference, WAIM, Jiuzhaigou, China,* 471-482, New York, NY: Springer.

Magadza, T., Mukwazvure, A. & Supreethi, K. (2014). Exploring sentiment classification techniques in news articles. *International Journal of IT & Knowledge Management (IJITKM)*, *8* (1), 55-58. Retrieved from <https://www.academia.edu> . Last accessed on 01/10/2019.

Mejova, Y. (2012). *Sentiment analysis within and across social media streams* (Doctoral dissertation). The University of Iowa. Retrieved from <https://www.academia.edu> . Last accessed on 01/10/2019.

Miner, G, Elder, J., Hill, T., Nisbet, R., Delen, D., & Fast, A. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Oxford, UK: Academic Press.

Moreo, A., Romero, M., Castro, J., Zurita, J. (2012). Lexicon-based comments-oriented news sentiment analyzer system. *Expert Systems with Applications*, *39* (10), 9166-9180. Retrieved from <https://www.sciencedirect.com> . Last accessed on 01/10/2019.

Mudinas, A. (2012). *EmotionLookupTable.txt*. Retrieved from <https://github.com/AndMu/Wikiled.Sentiment/tree/master/src/Resources/Library/Standard>
Last accessed on 01/10/2019.

Shaikh, M., Prendinger, H., & Ishizuka, M. (2008). Sentiment assessment of text by analyzing linguistic features and contextual valence assignment. *Applied Artificial Intelligence*, *22*, 558–601. Taylor & Francis Group.

Statista [Database website]. (2019). Newspapers ranked by circulation in the United Kingdom
    (UK) in 2019. Retrieved from <https://www.statista.com> . Last accessed on 01/10/2019.

Statista [Database website]. (2019). Leading newspapers in the U.S. 2017-2019, by circulation. Retrieved from <https://www.statista.com > . Last accessed on 01/10/2019.

Taboada, M. (2009). Comparable corpora for cross-linguistic sentiment analysis.
    *Conference of the American Association for Corpus Linguistics* (AACL 2009), Edmontn, Alberta. Retrieved from <https://www.ualberta.ca/~aacl2009/PDFs/Taboada2009AACL.pdf>.  Last accessed on 01/10/2019.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Journal of the Association for Computational Linguistics*, *37* (2), 267-307. Retrieved from <https://www.academia.edu>  . Last accessed on 01/10/2019.

The Sun [Online newspaper]. (2019). Retrieved from<https://www.thesun.co.uk>.
    Last accessed on 01/10/2019.

USA Today [Online newspaper]. (2019). Retrieved from https://www.usatoday.com>
    Last accessed on 01/10/2019.

Worldatlas [Database website]. (2019). The 10 most popular daily newspapers in the United States. Retrieved from <https://www.worldatlas.com>.
    Last accessed on 01/10/2019.