An Analysis of Online Twitter Sentiment Surrounding the European Refugee Crisis

David Pope and Josephine Griffith
College of Engineering and Informatics National University of Ireland Galway, Galway, Ireland

Keywords: Social Media Analysis, Sentiment Analysis, Refugee Crisis.

Abstract: Using existing natural language and sentiment analysis techniques, this study explores different dimensions of mood states of tweet content relating to the refugee crisis in Europe. The study has two main goals. The first goal is to compare the mood states of negative emotion, positive emotion, anger and anxiety across two populations (English and German speaking). The second goal is to discover if a link exists between significant real-world events relating to the refugee crisis and online sentiment on Twitter. Gaining an insight into this comparison and relationship can help us firstly, to better understand how these events shape public attitudes towards refugees and secondly, how online expressions of emotion are affected by significant events.

1 INTRODUCTION

Due to the rapid growth of online social media over the last decade, and the ease of obtaining social media data, many computing techniques within Information Retrieval and machine learning domains have been applied to social media data. Sentiment analysis techniques are particularly relevant due to the nature of social media data which often contains a diverse set of human opinions.

From a high-level perspective, the goal of the application of sentiment analysis (also known as opinion mining) to social media data is to analyse the opinions of online social media users with respect to a particular topic or subject matter. According to Tumasjan et al. (Tumasjan et al., 2010) there are now many streams and applications of sentiment analysis research, examples of which are in the areas of product marketing, project management and politics. There are two main approaches to perform sentiment analysis: a lexical-based approach in which large dictionaries of psychologically evaluated words, terms and word stems are used to calculate a number different mood states (Tumasjan et al., 2010); and a supervised machine learning approach using classification techniques such as Support Vector machines, Naive Bayes or Maximum Entropy to learn, given training data, the sentiment associated with some new unseen data (Thelwall et al., 2011).

In this paper, existing sentiment analysis techniques are applied to a twitter data collection relating to the current refugee crisis in Europe. This refugee crisis has been described by organisations such as the United Nations Refugee Agency (UNHCR) and the European Commission for Humanitarian aid and Civil Protection (ECHO) as the planet's worst refugee crisis since the second World War[1]. Sparked by violent and brutal civil war in the Middle East, millions of refugees have fled to Europe in seek of shelter and protection, particularly from Syria (UNS, 2016). This migration has been received with varying levels of emotions in Europe and it is these "varying levels of emotions" that are under study in this paper.

The Twitter dataset, gathered over 68 days from November 2015 to early January 2016, comprises of English-language and German-language tweets, each language representing roughly half of the dataset. The aim of the work is firstly to ascertain if there are any noticeable differences in sentiment across the English and German populations (as represented by the two tweet collections). In addition, during the collection of the data, two noticeable events occurred in Europe: the Paris attacks on 13th December 2015 and the Cologne attacks on the 31st December 2015. The sentiment expressed in both datasets from dates surrounding these two events will also be compared. The sentiment categories chosen for the comparison are: negative emotion, anger, anxiety and positive emotion. A lexical approach (using Linguistic Inquiry and Word Count, LIWC (Pennebaker et al., 2015b))

[1] Syria crisis: Echo factsheet (online) posted May 2016.

Pope, D. and Griffith, J.
An Analysis of Online Twitter Sentiment Surrounding the European Refugee Crisis.
DOI: 10.5220/0006051902990306
In *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) - Volume 1: KDIR*, pages 299-306
ISBN: 978-989-758-203-5

is taken to ensure that the findings are psychologically strong with respect to the sentiment categories under analysis.

To date, there has been little analysis of online sentiment surrounding the current refugee crisis in Europe. One recent study completed by Coletto et al. used a supervised machine learning approach to explore positive and negative polarized Twitter sentiment relating to the refugee crisis in Europe across time and space (Coletto et al., 2016). In contrast to Coletto et al.'s research, this study not only uses a different lexical approach to perform the sentiment analysis of tweets, and uses tweets in two languages, it also uses two additional mood (anger and anxiety). Thus the contributions of this work are in the gathering and sentiment analysis of both an English and German dataset relating to the refugee crisis in Europe and comparing the sentiment in these datasets across languages and events.

The paper outline is as follows: previous work relating to social media sentiment analysis is outlined in Section 2. The methodology is presented in Section 3 where the data gathering, data cleansing and LIWC categorisation of sentiment is described. Two sets of related results are presented: Section 4.1 presents results on general trends found across the two populations focusing on four sentiment categories; Section 4.2 focuses on two events in both populations while again considering the same four sentiment categories and also considering word maps of top-occurring terms on the days around the events.

2 RELATED WORK

The area of sentiment analysis or "opinion mining" has received increased interest in Computer Science and other disciplines over the last number of years. Sentiment analysis techniques have been used to solve a variety of problems. As Caragea et al. discuss (Caragea et al., 2014), using sentiment analysis techniques as forecasting or predictive tools can help researchers, business analysts, business leaders, disaster response personal, economists, politicians, journalists and many more to extract and categorize raw data from online social media and transform it into actionable knowledge. With this newly obtained and valuable knowledge then comes effective decision making in the problem domain, whatever domain it may be.

Examples include predicting future stock market prices by analysing the mood of Twitter users (Bollen et al., 2011) and predicting German political elections with greater accuracy than traditional opinion polls or surveys (Tumasjan et al., 2010). Asur et al. (Asur and Huberman, 2010) have also demonstrated the predictive power of social media in their study of forecasting box office revenues. They were successful at forecasting box office revenues for upcoming movie releases better than any leading market-based predictors.

Li et al. also demonstrated the power of sentiment analysis in their study exploring the relationship between mood and changes in the weather (Li et al., 2014). Results indicated that a relationship did exist between online sentiment on Twitter and changes in the weather, most notably during periods of increased snow depth. The results found that an increase in snow depth correlated with an increase in two dimensions of the profile of mood state scale, *Depression-Dejection* and *Anger-Hostility*.

Data from other social media platforms, such as Facebook, have also been used as a platform to perform sentiment analysis. For example Kramer carried out a sentiment analysis study exploring the gross national happiness of Facebook users (Kramer, 2010). Gilbert and Karahalios used the older social media platform LiveJournal to explore the relationship between widespread worry ion LiveJournal and fluctuations in stock market prices (Gilbert and Karahalios, 2009).

There are generally two main approaches taken in performing a sentiment analysis study and include:

- Lexical Approach which involves examining the semantic orientation of words and phrases in a piece of text in order to attempt to extract and calculate sentiment polarity. The lexical approach incorporates the use of a lexicon or a dictionary of pre-defined words, word stems or phrases that have been semantically tagged into a number of different sentiment categories. Each word and phrase in the lexicon is tagged with a score to signify mood intensity. Words, parts of words and word stems contained within the lexicon can also be categorized into other more explicit dimensions of mood states such as *anxiety* and *anger*. Tumasjan et al. used this approach to construct psychological profiles of election candidates in the 2011 German Federal election using LIWC (Tumasjan et al., 2010).
- Machine Learning approach where typically a supervised machine learning approach is used for sentiment analysis which requires a labelled set of training data. Human evaluators are tasked with labelling a random sample of data obtained from the corpus into the desired sentiment categories, i.e., negative, positive and neutral sentiment. The data produced by the human evaluators is used as training data for the chosen supervised classification algorithm to produce a model.

Algorithms such as Naive Bayes or Support Vector Machines are often used. The remaining data from the corpus is passed to the model to produce the sentiment results. Thelwall et al. (Thelwall et al., 2011) and Asure et al. (Asur and Huberman, 2010) used this approach in their studies.

3 METHODOLOGY

This section provides an overview of the methodology used to produce the results of both the sentiment analysis and term frequency stages. Firstly, English and German tweets were gathered from Twitter. All tweets gathered on a particular day are treated as a unit of analysis and a sentiment score is obtained for the tweets for each day. The tweet corpus is therefore analysed based on the creation/publish date and the language. Following this, the tweet corpus is passed through a number of data pre-processing and filtering techniques to prepare the data for the sentiment analysis stage. The tweet corpus is then passed to the LIWC tool which uses its own large internal lexicons to produce the sentiment results. Sentiment scores produced by the LIWC tool are recorded and stored per day for each day in the time range. Using the LIWC results the mean and standard deviations are calculated to establish the significance of the results obtained. Finally a term frequency (TF) process is performed for the days around two significant events to find the top terms, and potential topics, being discussed by Twitter users on those days.

3.1 Data Gathering and Data Pre-processing

Using the public Twitter REST search API, over 1.6 million tweets were gathered between the 6th November 2015 and 13th January 2016. All tweets gathered were added to a tweet corpus of which 902,139 were English tweets and 702,852 were German tweets. The basic keywords used are listed in Table 1 and are all variants of words used to describe the refugee crisis in English and the corresponding German translations. These keywords, and variations of the keywords including capitalisation and the use of the "#" character, are passed as parameter arguments to the Twitter search API for English and German tweets.

Tweets were organised according to the language used and the day of creation. All tweets per day, in each language, were treated as a unit for future processing. To help reduce the bias in the sentiment analysis results, tweet objects that contain identical unique identifiers and retweeted tweet objects were removed from the corpus. As the sentiment approach used is lexical, the repeating of words in retweets, which are in the sentiment dictionaries, would most likely give a much higher sentiment score thus skewing the results.

Table 1: List of English and German Search Terms used for Twitter Search API.

English Keywords	German Keywords
refugee	flüchtling
refugee crisis	flüchtlingskrise
migrant crisis	migrationskrise
economic migrant	wirtschaftsflüchtling
asylum seeker	asylbewerber

To help reduce the influence of *noisy* data on results, all tweets were passed through a number of custom filtering and data transformation techniques to standardise the terms and symbols used across all tweets. Hyperlinks and the characters "@" and "#" were removed. The character "&" was transformed to "and" or "und" depending on the language and "U.S.A" was transformed to the full name. This step was taken in accordance with instructions outlined in the LIWC documentation (Pennebaker et al., 2015b).

Following the data pre-processing stages, 53,355 completely unique tweets remained of which 28,866 tweets were flagged as English and 24,469 tweets were flagged as German.

3.2 Sentiment Analysis using a Lexical Approach

The Linguistic Inquiry and Word Count (LIWC) tool is a natural language text processing (NLTP) tool that is available for academic use (Pennebaker et al., 2015b) and has been used in previous studies of social media data to perform sentiment analysis (Tumasjan et al., 2010). LIWC uses a lexical approach to perform sentiment analysis and the large LIWC internal dictionaries, also known as lexicons, have been refined and developed over a number of years by the LIWC development team which includes psychologists and computer scientists (Pennebaker et al., 2015b). It is for this reason that a lexical approach, and specifically LIWC, was chosen for this study rather than a machine learning approach. With a machine learning approach, the results would be dependent on the quality of the training data. The training data in our approach would be difficult to obtain, given that the data is in both English and German and that human evaluators may not always be able to label consistently according to the sentiment categories used in the study. Although a number of sentiment categories exist in LIWC, four LIWC categories were selected for the sentiment analysis results presented in this paper and are listed here along with the words and word stems (represented by “*”) which are associated with those sentiment categories:

- **Positive Emotion:** words such as *love, nice, sweet, fantastic, heal, decent, honest, hope*, word stems such as *ecsta* (ecstatic), encourag* (encourage), magnific* (magnificent)*, and emoticons associated with positive emotion such as *:), (:*.
- **Negative Emotion:** words and word stems such as *agony, destruct, pain, resent, ignorant, dissatisf* (dissatisfaction), outrag* (outrage), vulnerab* (vulnerable)* and *suffering*, and emoticons associated with negative emotions such as :(are also included in the negative emotion category.
- **Anger:** words and word stems such as *hate, kill, brutal, hostil* (hostile), rude, sinister, rape, prejudic* (prejudice), beaten*, *aggressive*.
- **Anxiety:** words and word stems such as *worried, feared, nervous, worry, anxious, afraid, embarras* (embarrassed), paranoi* (paranoid), suspico* (suspicious).*

LIWC also contains a mean score for each of the above listed categories for both English and German languages. These means are available for different data sources, one of which is Twitter for the English language (Pennebaker et al., 2015b) and general data for the German language. We use these means as a baseline to compare the sentiment scores of each experiment as outlined in the LIWC 2015 and 2007 documentation guides (Pennebaker et al., 2015a) (Pennebaker et al., 2007). We will refer to these means as “LIWC mean” in the results. In addition, we also calculate our own mean score based only on our dataset. We refer to this mean as “Calculated mean” in the results.

The tweet object creation date was used to order all tweet objects resulting from the pre-processing stage. For each language, and for each day, the tweet text contained within each tweet object for that day was extracted and stored in a single file. Each text file representing each day and language was then passed to the LIWC tool. A category score for each of the LIWC categories was calculated per day, per language. LIWC generates this score based on the total number of words present within the tweets that match words, word stems, emoticons and expressions categorized within the specified categories of the English and German internal LIWC dictionaries.

In addition to using the LIWC scores for the four emotion categories, the analysis of two significant events is supported by the generation of weighted word maps. These were produced using the top-30 words (with stop words removed) found per day and graphically represented using the wordclouds tool[2].

4 RESULTS AND DISCUSSION

Section 4.1 presents the LIWC results focusing on the comparison across sentiment categories and populations. Section 4.2 also considers the same LIWC results, in addition to word maps, but focuses on only two days in particular: 13th September 2015 and 31st December 2015.

For all sentiment category graphs showing LIWC scores over the time range of the study, the *X* axis represents the day in the time range and the *Y* axis contains the daily scores produced by LIWC for a particular emotion category. Each sentiment category graph contains two means per language: the LIWC mean is the average LIWC mean (independent of the refugee crisis data) and is unique for each category as already discussed. The “calculated” mean represents the calculated average score using the results produced by the LIWC tool for the refugee crisis data gathered in this study. Using this “calculated” mean the standard deviation was also calculated for each category to allow the significance of the LIWC results for each day to be discussed. As is evident in the sentiment category graphs in the results section, the results fluctuate, over time, for each day, demonstrating, for example, an increase or decrease in a sentiment category.

4.1 Sentiment Category Results: per Population and per Day

This section will present and discuss the LIWC results in detail comparing both English and German results together, per sentiment category: Negative Emotion, Anger, Anxiety and Positive Emotion. Recall that each tweet contains at least one of the search terms relating to the refugee crisis, as outlined in Section 3.1.

- **Negative Emotion.** The LIWC results for negative emotion are illustrated in Figure 1 for both English and German. The calculated means for the English and German sets of results are both above the LIWC mean. This suggests that, over the entire datasets, for each day and in each language, tweets relating to the refugee crisis show a higher negative emotion level compared to the LIWC averages. It can be seen in Figure 1 that there are very few days that are below the LIWC negative emotion mean. Also notable is the lack

[2]Free online wordcloud generator.

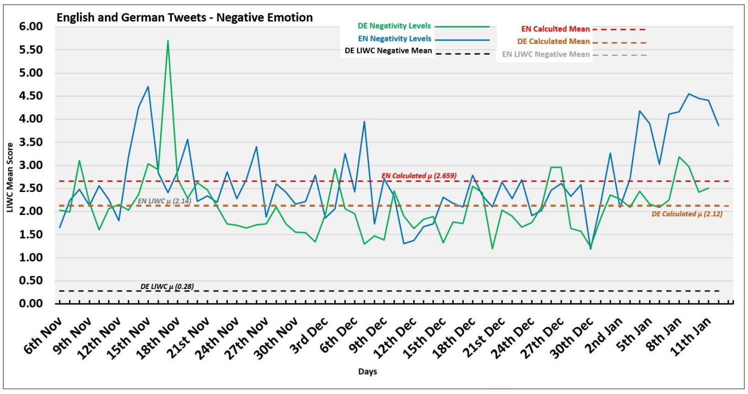

Figure 1: English and German LIWC Negative Emotion.

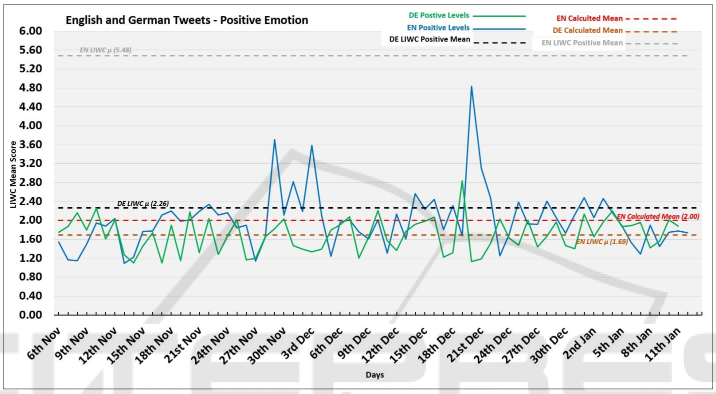

Figure 2: English and German LIWC Positive Emotion Levels.

of increased negative emotion after the 31st December in the German tweet data compared to the English tweet data. While there is an increase in expressed negative emotion in both English and German tweet datasets after the 13th November, there is a difference in the significance of this increase. For the English tweet data, there is an increase of 2.5 standard deviations above the mean after the 13th November whereas for the German tweet data there is a sharper increase to over 5 standard deviations above the calculated mean. However, in both cases there is a sharp decrease in negative emotion following the increases.

- **Positive Emotion.** The LIWC results for positive emotion are illustrated in Figure 2 for both the English and German datasets. Both calculated means in Figure 2 are much lower than the LIWC means. In fact positive emotion only surpasses the LIWC mean for one day for both languages. Thus, this suggests that across both result sets, for English and German tweets, there is low positive emotion expressed by Twitter users. This pattern does correlate to the high negative emotion levels expressed in Figure 1 previously. The positive emotion levels in Figure 2 for the German tweet data illustrate a much more aggressive fluctuation in scores across the time range under analysis.

- **Anger.** A similar pattern to that seen for negative emotion levels can be seen in Figure 3 where the sentiment scores for *anger* are mostly higher than the LIWC means. This is highlighted in Figure 3 where the English tweet anger levels only drop below the LIWC mean for one of the 68 days. The German tweet anger levels indicate that there were also a smaller number of days below the LIWC mean. Across both the English and German tweet datasets, a similar fluctuation pattern in anger levels over the entire time range is visible.

- **Anxiety.** The calculated mean values for anxiety levels for the English and German tweet datasets are represented in Figure 4 where it can be seen that both calculated means are above the LIWC means. This suggests that across the entire dataset, anxiety levels are higher than the expected LIWC mean. This was also seen with negative emotion and with anger levels. It is clear from Figure 4 that for the English tweet dataset, anxiety levels drop below the LIWC anxiety mean

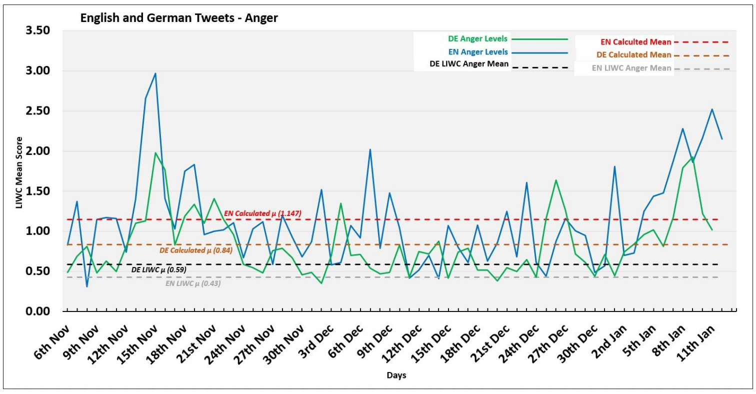

Figure 3: English and German LIWC Anger.

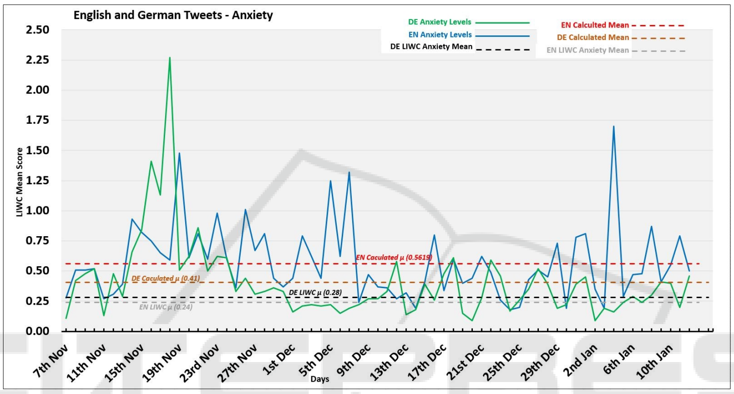

Figure 4: English and German LIWC Anxiety.

for only 5 of the 68 days within the time range under analysis. Comparing the English and German graphs in Figure 4 it can be see that, in contrast to German tweet anxiety levels, English tweet anxiety levels show a more aggressive fluctuation in scores. In comparison, for the German tweets, a period of relevant "calmness" following a period of sharp increases and decreases of anxiety levels between the 13th and 26th November, can be seen.

4.2 Sentiment Category Results per Event: Paris and Cologne Attacks

This section focuses on two events in particular: the days after the 13th November 2015 and the days after the 31st December 2015 (following the night of the Cologne attacks in Germany). In order to better establish if Twitter users are discussing the events that took place in Paris and Cologne, weighted word maps, produced by calculating the frequency of the top-30 terms, are generated. These weighted word maps are shown in Figures 5 and 6 and illustrate that in both the English and German result sets, the Paris attacks feature as a top term in tweets published after the event on the 13th November 2015. The top German terms include "refugee coordinator" (flüchtlingskoordinator), "federal government" (bundesregierung) and "paris".

References to the events that occurred in Cologne do not begin to surface in the English top terms until the 7th January, almost a week after the event occurred. This is also the case for the German dataset where references to the Cologne attacks do not surface until the 2nd January and only begin to appear in the top terms on the 4th January 2016. The list of top German terms include "dänemark" (denmark), "deutschland" (germany), "cologne" (köln), "leads" (führt), "passport control" (passkontrollen), "deported" (abgeschoben) and "shots" (schüsse). This is an interesting finding, as it was reported that there was a delay in the reporting of the events in Cologne (Scally, 2016).

With respect to the sentiment category of *Negative Emotion*, perhaps what is most obvious in Figure 1 is the increase in *Negative Emotion* after 13th November, the night of the Paris terrorist attacks. The English tweet results in Figure 1 show an increase of 2.5 standard deviations above the calculated mean and the

Figure 5: Top English Terms on 14th November 2015 and Top German Terms on 15th November 2015.

Figure 6: Top English Terms on 7th January 2016 and Top German Terms on 4th January 2016.

German tweet results show negativity levels of 5 standard deviations above the calculated mean after the 13th November. However, already mentioned, a sharp decrease in negative emotion quickly follows the increases after 13th November for both the English and German datasets. What is also interesting in Figure 1 is the difference in negative emotion after the 31st December 2015. For the English tweets, we can see a gradual increase in negative emotion (up to 2.5 standard deviations above the calculated mean), whereas there is little change in negative emotion levels in the German tweets.

With respect to the sentiment category of *Anger*, Figure 3 clearly illustrates an increase in anger after the 13th November and after the 31st January for both German and English results, where in both cases, scores reach 3 standard deviations above the calculated means.

With respect to the sentiment category of *anxiety*, as shown in Figure 4, there is a prominent increase in anxiety levels in German tweets after the 13th November 2015 reaching a high of over 5 standard deviations above the calculated mean. In contrast to German anxiety levels, English Tweet anxiety levels display a more aggressive fluctuation with an increase in anxiety levels after the 13th November 2015 in the English tweet dataset. These higher peaks of anxiety in the English tweets appear to decrease and increase in intensity over a number of days and weeks after the 13th November, finally decreasing to the LIWC mean and calculated mean after the 7th December. There is evidence of a period of heightened anxiety in English tweets between the 13th November and the 7th December 2015. It is also interesting to note a level of “calmness” in the German tweet scores after the 25th November where there is no obvious or significant increase or decrease in anxiety levels, even after the Cologne attacks on 31st December 2015.

5 CONCLUSION AND FUTURE WORK

Using existing sentiment analysis techniques, the goal of this study was to explore different mood states of tweet content relating to the refugee crisis in Europe. In addition, the goal of identifying changes in online expressions of emotion in tweet content triggered by significant offline events relating to the refugee crisis was also undertaken in this study.

The sentiment categories of negative emotion, positive emotion, anger and anxiety were analysed across two populations (English and German speaking) and across 68 days. Two significant events occurred during these 68 days (the Paris Terrorist attacks and the Cologne attacks) and these events were analysed by considering the four sentiment categories in addition to the frequency of words used in tweets around those days. A lexical approach using the LIWC tool was adopted which, in addition to producing sentiment scores per category and per language, provided mean scores for each category.

The two main goals of this study were achieved. Firstly the results from the sentiment analysis stage show some interesting trends and commonalities across languages and days. Coupling these sentiment analysis results with the results of the term frequency analysis the second goal of determining if online twitter sentiment is affected by significant events relating to the refugee crisis was also achieved. The results of this study show interesting trends and commonalities across languages and days. However, it is important that we do not base conclusions about refugees and the refugee crisis in Europe on subjective opinions. Instead, using existing sentiment analysis techniques the results of this study may help us to better understand how online expressions of emotion and attitudes towards refugees are impacted by significant offline events such as the Paris and Cologne attacks.

Other LIWC categories will be explored in future work. It is evident in many of the sentiment graph results that there are a number of other spikes of heightened expressed emotion. In order to establish the topics of discussion on these days the researchers are currently looking into topic detection using unsupervised machine learning clustering techniques.

REFERENCES

(2016). Syria regional refugee response inter-agency information sharing portal. [Online Website: Accessed: 2016-06-15].

Asur, S. and Huberman, B. (2010). Predicting the future with social media. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference*.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*.

Caragea, C., Squicciarini, A., Stehle, S., Neppalli, K., and Tapia, A. (2014). Mapping moods: Geo-mapped sentiment analysis during hurricane sandy. *11th International Conference on Information Systems for Crisis Response and Management*.

Coletto, M., Lucchese, C., Muntean, C., Nardini, F. M., Esuli, A., Renso, C., and Perego, R. (May 2016). Sentiment-enhanced multidimensional analysis of online social networks: Perception of the mediterranean refugees crisis. *Under Review*.

Gilbert, E. and Karahalios, K. (2009). Widespread worry and the stock market. *4th International AAAI Conference on Weblogs and Social Media*.

Kramer, A. (2010). An unobtrusive behavioral model of "gross national happiness". *Proc. of SIG CHI Conference on Human Factors in Computing Systems*.

Li, J., Wang, X., and Hovy, E. (2014). What a nasty day: Exploring mood-weather relationship from twitter. *Proc. of 23rd ACM International Conference on Conference on Information and Knowledge Management*.

Pennebaker, J., Boyd, R., Jordan, K., and Blackburn, K. (August 2015a). The development and psychometric properties of liwc2015'.

Pennebaker, J., Chung, C., Ireland, M., Gonzales, A., Booth, R., and Francis, M. (2007). The development and psychometric properties of liwc2007.

Pennebaker, J., Roger, J., Booth, R., Boyd, L., and Francis, M. (August 2015b). Linguistic inquiry and word count: Liwc2015, operator manual.

Scally, D. (2016). Media under scrutiny over slow response to cologne attacks. *The Irish Times Newspaper*.

Thelwall, M., Buckley, K., and Paltoglou, G. (2011). Sentiment in twitter events. *Journal of the American Society for Information Science and Technology*, 62.

Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *4th International AAAI Conference on Weblogs and Social Media*.