Title	A Twitter sentiment gold standard for the Brexit referendum
Author(s)	Hürlimann, Manuela; Davis, Brian; Cortis, Keith; Freitas, André; Handschuh, Siegfried; Fernández, Sergio
Publication Date	2016-09-12
Publication Information	Manuela Hürlimann, Brian Davis, Keith Cortis, André Freitas, Siegfried Handschuh, Sergio Fernández (2016). A Twitter Sentiment Gold Standard for the Brexit Referendum. Paper presented at the Proceedings of the 12th International Conference on Semantic Systems, Leipzig, Germany.
Publisher	CEUR Workshop Proceedings
Link to publisher's version	http://ceur-ws.org/Vol-1695/paper26.pdf
Item record	http://hdl.handle.net/10379/6241

A Twitter Sentiment Gold Standard for the Brexit Referendum

Manuela Hürlimann,
Brian Davis
Insight Centre for Data Analytics
National University of Ireland
Galway, Ireland
{first.last}@insight-centre.org

Keith Cortis,
André Freitas,
Siegfried Handschuh
University of Passau
Germany
{first.last}@uni-passau.de

Sergio Fernández
Redlink GmbH
Salzburg
Austria
sergio.fernandez@redlink.co

ABSTRACT
In this paper, we present a sentiment-annotated Twitter gold standard for the Brexit referendum. The data set consists of 2,000 Twitter messages ("tweets") annotated with information about the sentiment expressed, the strength of the sentiment, and context dependence. This is a valuable resource for social media-based opinion mining in the context of political events.

CCS Concepts
• **Computing methodologies→Artificial intelligence→Natural language processing→Language resources.**

Keywords
Sentiment Analysis; Political Opinion Mining; Twitter; gold standard; annotation.

1. INTRODUCTION
Popular referenda provide a rich setting for understanding the social and discourse dynamics behind a focussed political discussion. Under these settings, opinion mining and sentiment analysis over social media are fundamental tools to provide systematic prospective and retrospective insights, supporting an analysis of the underlying political processes and dynamics.

However, referendum-type events require the application of different techniques and resources for opinion mining and sentiment analysis as these events have a distinctive social dynamics and political discourse. The availability of language resources to ground the discourse analysis, and the construction of supervised classification methods play a fundamental role for

pushing forward our ability to build systems which can support the interpretation of social media discourse.

Aiming to support the evolution of the classification methods, this paper presents a dataset of sentiment-annotated social posts targeting the historical event of the United Kingdom European Union membership referendum[1] ("Brexit"), which took place on June 23, 2016. Data collection and annotation were carried out in the context of the SSIX H2020 project [2], which targets the creation of customisable sentiment metrics for social media. The dataset, containing 2,000 annotated tweets, was collected from Twitter prior to the event. The dataset has been published[2] under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) licence[3], for general use.

2. RELATED WORK
A number of datasets have been created in the context of Brexit. The #ImagineEurope project[4] [3, 4] collected a dataset of tweets using a set of hashtags that are related to the Brexit referendum, such as the ones used by the leave and remain campaigns. Examples of hashtags include *#migrant*, *#refugee*, *#strongerin*, *#leadnotleave*. The collection of this dataset was initiated several months prior to the referendum (7th August 2015).

Priego [8, 9] collected datasets of tweets from the official campaign accounts over a period of nine days (June 12 to June 21, 2016). While the "Vote Leave" set [8] contains 1,100 Tweets by *@vote_leave*, the "Stronger In" set [9] consists of 1,005 tweets by *@StrongerIn*.

A Twitter-based dataset with a wider scope is the Twitter political corpus [6]. It was collected during 2009 and consists of two subcorpora. The first is a randomly selected set of 2000 tweets from Twitter's "spritzer" feed collected between June 1, 2009 and Dec 31, 2009, while the tweets for the second corpus were randomly selected from a subset of tweets which contained at least one political keyword each. The aim of this study was to develop learning algorithms which link political statements on Twitter to general opinions about government and politicians. A

---

[1] http://www.legislation.gov.uk/ukpga/2015/36/crossheading/the-referendum

[2] https://bitbucket.org/ssix-project/brexit-gold-standard

[3] https://creativecommons.org/licenses/by-sa/4.0

[4] Part of the Economic and Social Research Council's UK in a Changing Europe programme; cp. http://ukandeu.ac.uk/

classifier was trained on these corpora for sentiment analysis, which was used for predicting presidential approval polls.

Several researchers have studied the social media debate surrounding Brexit. In [5], the authors presented a demonstrator which visualises the Twittersphere debate on whether the UK should remain in or leave the European Union. It shows the different discussion topics identified by the different search strategies of the collected data, namely hashtag search terms, extraction from the full stream and following specific users.

A study by [7] analysed over 1.5 million tweets related to the referendum up until 13 May 2016. Tweets were identified by common referendum-related terms, such as #brexit; #StrongerIn; "uk eu vote", and then analysed for the use of hashtags, citations, and general sentiment for leaving or remaining in the EU. Results from the study suggest that the Brexit referendum is very close, with a clear preference among Twitter users for leaving the EU.

The real-time monitoring of the Brexit campaign set up by [10] showed indications that 3 hours before the polls closed in the UK, the split between *#VoteRemain* and *#VoteLeave* was roughly 40:60 for the last four hours of voting. This study also dealt with parallel investigations and analyses of how the market was reacting and how currencies were changing, which are of real, actionable value to financial firms including hedge funds, government bodies, politicians, and policy makers.

3. METHOD
3.1. Sampling and Filtering
In order to collect the data set, we sampled 2000 tweets uniformly at random from a Twitter stream which was set up to track 75 keywords, including hashtags and account names. Criteria for chosen keywords were based on the manual identification of common keywords associated with content relevant to Brexit, for example #eureferendum, #votein, #voteleave. *Appendix A* provides a list of Twitter tracking keywords (hashtags and Twitter handles). Data collection on this stream between May 4 and May 6, 2016 (inclusive) resulted in a population of 149,331 tweets. Before sampling, filters were applied to exclude spam and irrelevant content: for example, we discarded very short contents (less than three characters) and users with suspiciously high activity, (i.e. more than 100 tweets per day, which should exclude most spammers but include prolific real tweeters). Furthermore, only tweets published between 6am and 11pm GMT were considered in order to increase coverage of European postings. These measures reduced the population for sampling to 20,104 tweets.

3.2. Annotation
The 2000 tweets thus sampled were presented to three annotators, all proficient in English, who created the following annotations for each tweet.

Sentiment: assign one of the following (or leave blank if cannot decide):

- Stay: the tweet is in favour of the UK remaining in the EU
- Leave: the tweet is in favour of the UK leaving the EU

- Undecided: the tweet expresses indecision about the EU referendum
- Don't care/no sentiment: the tweeter does not care about the outcome of the referendum, or there is no sentiment expressed
- Irrelevant: the tweet is not about Brexit

Strength (only for tweets classified as "stay" or "leave"):

An integer between 1 (very weak) and 5 (very strong) expressing the strength of the "stay" or "leave" sentiment

Contextual dependency: one of the following[5]

- 0: interpretation of sentiment in tweet does not depend on external sources
- 1: interpretation of sentiment in tweet depends on external sources (e.g. articles or images that are linked)

The five opinion categories and strength annotations support a fine-grained view on the opinion landscape. Furthermore, the contextual dependency option provides an indication of the difficulty of scoring a tweet, which is a fundamental feature for the construction of opinion mining classifiers.

3.3. Agreement and Consolidation
In Table 1 we present the standard inter-rater agreement metrics for each of the annotations.

Table 1: Inter-rater metrics for each annotation type

	Fleiss' Kappa	Observed Agreement
Sentiment	0.394	0.537
Strength	0.240	0.439
Context	0.156	0.748

We achieve moderate Fleiss agreement for sentiment and strength, and fair agreement for context. Average observed agreement gives an indication of the difficulty of this annotation task. The strength assignment is the most difficult, while context dependency is relatively straightforward to determine.

Table 2 below shows the distribution of tweets with regard to the number of annotators who agreed on its opinion annotation, providing a different view of agreement.

Table 2: Agreement on sentiment

	# tweets	%
Unanimous	785	39.25
Two opinions	870	43.5
Three opinions	345	17.25
Total	2000	100

---

[5] For the contextual dependency annotation, annotators were advised to follow links where necessary for their decision.

We base our consolidation procedure on the three categories of agreement presented in Table 2. We use a) a majority vote for the opinion and contextual dependency annotations and b) the average (rounded to the nearest integer) for the strength annotation for the first two rows (unanimous and two different opinions). Cases where three different options were selected by the annotators were consolidated manually by a fourth person who had not previously been involved in the annotation. For further information on the resulting data set, cp. Section 4 below.

4. DATA SET DESCRIPTION

The gold standard obtained according to the method described in Section 3 above consists of a total of 2000 tweets. The distribution of sentiment annotations can be seen in Table 3.

Table 3: Distribution of sentiment annotations

Annotation	# tweets	%
stay	430	21.5
leave	816	40.8
undecided	82	4.1
no sentiment/don't care	502	25.1
irrelevant	148	7.4
[left blank]	22	1.1

The great number of "leave" tweets in our data set reflects the overall tweeting behaviour, as also identified by [7]. Very few tweets display an "undecided" sentiment, in line with observations by [1] that strong opinions predominate on Twitter. Our data display a similar bi-modal distribution if we only consider "stay", "leave" and "undecided". Note that the distribution of sentiment labels in our gold standard does not make any statement about the referendum outcome.

An example of a strong "stay" tweet is given in Example (1a) and a strong "leave" tweet in Example (1b).

(1a) *@GeorgeKaburu LOL - how might you think that? Against Brexit is an UNDERSTATEMENT I believe it would be disastrous for the UK and Europe.*

(1b) *@David_Cameron aren't you listening WE DON'T WANT THESE PEOPLE IN UK #VoteLeave @vote_leave @NoThanksEU @leavehq https://t.co/wNmrFqQNFW*

The low percentage of irrelevant tweets shows the usefulness of our tracking keywords in retrieving content which is relevant to Brexit. Many of these irrelevant tweets are in languages other than English. Some hashtags are used ambiguously, such as in Example (2), where "#takecontrol" is used in the context of yoga rather than the "Leave" campaign, who coined the phrase.

(2) *#yoga #positivethinking #letgo #accept #learn #grow #takecontrol #workhard #workout #bebetter #wisdom #loveyourself https://t.co/cDmjj4mhbG*

A rather large number of tweets were classified as not displaying any sentiment. These include tweets asking questions on the impact of Brexit (as in Example (3a)) or expressing non-opinionated interest in the issue (cp. Example (3b)).

(3a) *Will the rights to travel, live and work across the #EU change if there's a #Brexit? https://t.co/PJSu1IGQg1 #ukemplaw*

(3b) *Eagerly anticipating our #EUreferendum debate later today!*

Table 4 shows the distribution of opinion strength annotations in our data. We can see that there is a greater tendency for "leave" tweets to display strong opinions, while the "stay" opinions tend towards the weaker end of the scale. Both opinions, however, span the entire continuum.

Table 4: Distribution of strength annotations

Strength	#stay	%	# leave	%
1	88	20.5	131	16.1
2	193	44.9	304	37.3
3	135	31.4	263	32.2
4	13	3.0	112	13.7
5	1	0.2	6	0.7

Finally, 268 tweets (13.4%) were annotated as depending on context, while 1732 tweets (86.6%) were annotated as not depending on context. Table 5 shows a breakdown of context dependence by opinion annotation. We can see that the percentage of context-dependent tweets is rather stable across all opinion categories.

5. CONCLUSIONS

From a discourse perspective, the Brexit Twitter Sentiment Gold Standard provides a resource for observing the social and discourse dynamics behind the referendum (in contrast to most political corpora which will have as core discourse targets politicians and parties). The majority of the discourse acts present in the corpora can be categorised into 5 main classes: Event announcement (announcement of political events), linked fact-reference (link to larger factoid textual references), direct fact reference (summarised facts within the tweet), reference to political actors (containing opinions about the main political agents behind the opposing views) and informal statements (humorous or hate references).

Despite the fact that there are short dialogues (replies), the corpus does not contain complex instances of argumentation flows.

A limitation of social media-based analysis studies is that these present only a selective portion of society, since not everyone uses social media. These services are used predominantly by young and politically active people or by individuals with strong political views [1, 3]. This could be easily reflected in the Brexit results, where the majority of younger generation (age 18-44) voted to remain as opposed to people over age 45[6]. Such a result falls in line with the latest United Kingdom social media statistics, such as for Twitter were 72% of the users are between the age of

---

15-44[7], whilst for Facebook the most popular age group is 25-34[8] (26% of users).

Table 5: Distribution of context dependence annotations

Annotation	# context-dependent (%)	# context-independent (%)	total # tweets
stay	59 (13.7%)	371 (86.3%)	430
leave	134 (16.4%)	682 (83.6%)	816
undecided	15 (18.3%)	67 (81.7%)	82
no sentiment/ don't care	47 (9.4%)	455 (90.6%)	502
irrelevant	11 (7.4%)	137 (92.6%)	148
[left blank]	2 (9.1%)	20 (90.9%)	22
total	268 (13.4%)	1732 (86.6%)	2000

6. ACKNOWLEDGMENTS

7. REFERENCES

1. Barberá, P. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis, 23(1)*, 76-91.

2. Brian Davis, Keith Cortis, Laurentiu Vasiliu, Adamantios Koumpis, Ross McDermott, and Siegfried Handschuh. "Social sentiment indices powered by X-scores." In 2nd International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2016).

3. ImagineEurope. 2015a. Building a Twitter Dataset to Find out How People View the EU. https://storify.com/ImagineEurope/building-a-twitter-dataset-to-find-out-views-on-th. [Accessed 21 June 2016].

4. ImagineEurope. 2015b. EU Twitter Sentiment Analysis. Analysis of the sentiment in the Twittersphere towards the UK leaving or remaining in the EU. https://storify.com/ImagineEurope/initial-sentiment-analysis. [Accessed 21 June 2016].

5. Clare Llewellyn, Laura Cram. Brexit? Analyzing Opinion on the UK-EU Referendum within Twitter. In the Tenth International AAAI Conference on Web and Social Media. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13119

6. Micol Marchetti-Bowick and Nathanael Chambers. 2012. Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter. In *Proceedings of the European Association for Computational Linguistics*. Avignon, France. 2012. [Accessed 21 June 2016].

7. Ontotext. 2016. #BRexit Twitter Analysis. http://www.ontotext.com/documents/white_papers/brexit-twitter-analysis.pdf

8. Ernesto Priego. 2016a. "Vote Leave": Looking Into a Sample Archive of 1,100 vote_leave Tweets". Available from: https://epriego.wordpress.com/2016/06/21/vote-leave-looking-into-a-sample-archive-of-1100-vote_leave-tweets/. [Accessed 21 June 2016].

9. Ernesto Priego. 2016b. "Stronger In": Looking Into a Sample Archive of 1,005 StrongerIn Tweets. 21 June 2016. Available from: https://epriego.wordpress.com/2016/06/21/stronger-in-looking-into-a-sample-archive-of-1005-strongerin-tweets/. [Accessed 21 June 2016].

10. Amit Sheth. 2016. #Brexit: "there is a big trouble for #remain" — Some Lessons from Real-time #socialmedia Analysis. https://www.linkedin.com/pulse/brexit-big-trouble-remain-some-lessons-from-real-time-amit-sheth [Accessed 30 June 2016].

APPENDIX A: Twitter TRACKING KEYWORDS

brexit, #beleave, #betteroffin, #betteroffout, #Bremain, #Brexit, #brexitfears, #britainout, #britin, #EdEUref, #eukay, #eunegotiation, #EUpol, #EUpoll, #euref, #eureferendum, #eureform, #eurenegotiation, #europeanunion, #fudgeoff, #grassrootsout, #greenerin, #ImagineEurope, #LabourIn, #leadnotleave, #leafchaos, #leaveeu, #loveeuropeleaveeu, #MyImageOfTheEU, #no2eu, #notoeu, #projectfact, #projectfear, #ref, #referendum, #remain, #remaineu, #saferbritain, #StrongerIn, #theinvisableman, #theknoweu, #UKandEU, #ukineu, #UKRef, #UKreferendum, #votein, #voteleave, #voteout, #voteremain, #wrongthenwrongnow, #yes2eu, #yestoeu, #Davidcameron, #INtogether, #TakeControl, #euin , #euout, #NoEu, @vote_leave, @Vote_LeaveMedia, @StrongerIn, @StrongerInPress, @britinfluence, @lsebrexitvote, @eureferendum, @LeaveEUOfficial, @whatukthinks, @JuneExit, @EUinEUout, @Grassroots_out, @euromove, @UKandEU, @sayyes2europe, @Choice4Britain, @BrexitWatch

---