Online Supplementary Appendix: Once More, with Feeling: Using Sentiment Analysis to Improve Models of Relationships Between Non-State Actors*

Kevin T. Greene Caleb Lucas

October 10, 2019

*Kevin T. Greene is a Graduate Student in the Department of Political Science, University of Pittsburgh (ktg19@pitt.edu). Caleb Lucas is a Graduate Student in the Department of Political Science, Michigan State University (clucas@msu.edu).

Contents

A Data

Hezbollah began publishing an official weekly newspaper in 1984 (Khatib et al., 2014). Originally called “al-Ahed” (The Promise), the newspaper was renamed al-Intiqad (The Criticism) in 2001. Five years later, Hezbollah relaunched al-Ahed as a website hosting news published in al-Intiqad and general interest documents/reports produced by the group in English (Alagha, 2006). To construct our original corpus, we collect data from two sections of al-Ahed’s website; *Secretary General* and *Hezbollah Statements*.[1] The former are speeches delivered regularly by Hezbollah’s secretary general, Hassan Nasrallah, as well as other information about Nasrallah’s activities and his commentary on current events. Since Nasrallah essentially lives in hiding, his ability to directly communicate with supporters is largely limited to these speeches and interviews with journalists (Khatib et al., 2014).[2] As such, they are strategic opportunities to deliver substantive messages on important topics. The *Hezbollah Statements* section generally consists of shorter press releases about current events, as well as official documents/reports authored by the group’s propaganda wing. These official documents are originally in Arabic, but Hezbollah translates them into English. Our data covers the years 2006-2016 (all the years available when the data were collected). The contents of these pages were scraped and output to text files using the Python package beautifulsoup4.

[1]Since collecting the data, the website has changed and no longer publicly links to *Hezbollah Statements*. However, the links we scraped are live and still hosted on al-Ahed’s server at the time of writing.

[2]Nasrallah began limiting his public appearances in 2006, the year in which our corpus begins, due to fears for his physical safety.

A.1 Text Preprocessing

Since the data is scraped from a news website, it requires preprocessing (or cleaning) before analysis. We begin this step by dropping the date found at the beginning of most articles, the author(s) of the article if they appear outside the body of the article, and attribution of the source.[3] These strings are dropped since they are not a part of the actual article and could confuse the parser (nonsensical strings would result from these pieces of text being added to sentences in the article's body). We then standardize the names of the actors in our analysis such that multiple ways of referring to them are replaced with the same word (i.e. {al Qaida, Al Qaeda, al-Qaeda, alQaeda} = Qaeda). These alternate spellings are the result of transliterating their Arabic names into English slightly differently. Amal and Hamas are spelled uniformly.

Next, we deal with punctuation. All special characters are removed except for commas, semicolons, and end marks (this ensures we break between sentences correctly). However, we do not drop apostrophes if they occur in a contraction. This eliminates apostrophes that occur erroneously, a somewhat common occurrence in the corpus, but ensures the parser correctly identifies contractions. We retain numeric characters while parsing, but drop them before classifying since they are unhelpful features. Our corpus employs parentheses in an unusual way, sometimes containing unrelated information or even exclusively punctuation. Such instances can confuse the parser and do not offer substantive information to our measures. Parentheses are also used to define or expound upon a transliterated Arabic word. These strings are not helpful even in this case. As such, we drop all parentheses and the strings contained within them. Displayed below are four sentences from our corpus that use parentheses in these ways that demonstrate why we chose to perform this step.

---

[3]The source is typically a part of Hezbollah's organization and reads something like: "Source: Moqawama.org, 16-8-2007" or "Source: Hizbullah Media Relations."

Examples of Parentheses in the Corpus

1. This did not come at random; this was all planned for (...).

2. Even the great prophets like Allah's Messenger Muhammed (Peace be upon him and His Household), Prophet Essa (Jesus Christ; Pbuh), Prophet Moussa (Moses; Pbuh), and other prophets have been offended.

3. On another note, the leaderships congratulated Lebanese and Muslims on the Hijri New Year (Islamic New Year), praying that the occasion would unite Lebanese people against "Israel."

We then select sentences that contain the standardized name of an actor of interest or words that specifically refer to it. To find these words, we use Stanford's CoreNLP language software to parse our corpus and perform coreference resolution with their deterministic system (Manning et al., 2014; Recasens et al., 2013; Lee et al., 2011a; Raghunathan et al., 2010). See the coreference resolution section for more information on this process.

The sentences identified as referring to the actors in our analysis are then fed into scikit-learn's TfidfVectorizer to create document-term matrices.[4] Next, they are broken up into both unigrams (single words) and bigrams (two consecutive words). The elements of the matrices are the term frequency–inverse document frequency (tf-idf) adjusted word counts. This helps to minimize the effect of terms that appear frequently across many documents, consequently providing less information than more unique words.

[4]An alternative method would be to use aspect-based sentiment analysis (Liu, 2015). However, we classify the sentiment of sentences because we are interested in general alliances and rivalries and use other methods to investigate the content of those sentences. We encourage future researchers to pursue this approach.

A.2 Coreference Resolution

Coreference resolution, sometimes referred to as anaphora resolution, "is the process of determining whether two expressions in natural language refer to the same entity in the world" (Soon et al., 2001). Consider the simple example given by Lee et al. (2011b):

"[I] voted for [Nader] because [he] was most aligned with [my] values," [she] said.

In the example above, [I], [my], and [she] refer to the same entity, while [Nader] and [he] refer to another. Coreference resolution seeks to find when such associations occur (whether in a single sentence or or across a document) and then indexes them together. We use the Java implementation of Stanford's CoreNLP deterministic system (Manning et al., 2014; Recasens et al., 2013; Lee et al., 2011a; Raghunathan et al., 2010) to find coreferences and extract unique sentences in the chains of coreferences for each of the actors in our analysis. Using this method allows us to identify when Hezbollah refers directly to one of the actors in our analysis in a sentence, but refers again to them in one or more other sentences with one or more different words. For example, consider the following two sentences from one of Nasrallah's speech transcriptions in our corpus-

[Martyr Haj Imad] was a witness on the stage of the quantitative and qualitative development as well as on achievements and victories. In this context, His Eminence called for being faithful to the [martyrs] will and to the cause.

Coreference resolution allows us to identify that [martyrs] is referring to [Martyr Haj Imad], as the former is used to refer back to the latter. The example above demonstrates the importance of using coreference resolution in this project, as the text and sentiment in the latter sentence is substantively important and positive. If we merely selected sentences containing the exact name of each actor in our analysis, we would lose all of these coreferences and potentially bias our results.

A.3 Coding Process

The sentences used as training data for this article were coded separately by two individuals. Each person was given the same two sets of sentences, both of which were chosen randomly. The first set (n = 500) was sampled from sentences containing a reference to one of the non-state actors included in our analysis (Amal, Hamas, or al-Qaeda). The second set (n = 300) was sampled from sentences containing a reference to one of the anchor actors[5] (Israeli Prime Ministers or Imad Mughniyah). The coders then assigned the sentences in both sets a sentiment value of positive, negative, or neutral without consulting the other coder. After completing this, they met and reconciled values that were not the same across the two sets.

Sentiment values were assigned using the following criteria with examples from our corpus-

Negative Coding Criteria

1. The text referring to the actor is pejorative or condemnatory ("Qaeda terrorist group").
2. The text associates the actor with something clearly negative based on context ("Qaeda is, in fact, the other face of Daesh").

Neutral Coding Criteria

1. The text relays the actor's participation in some event in an impartial manner ("One day, there was a delegation from the Hamas leadership in Riyadh").
2. The text acknowledges the actor, but without judgment or association with something positive or negative ("We as well as Amal are political parties").

[5]See the Further Validation section for more information about how we validate our measure using two actors that have known relationships with Hezbollah that do not experience variation over time.

Positive Coding Criteria

1. The text uses words suggesting a positive relationship ("the brethren in Hamas").
2. The text referring to the actor uses general, religious or cultural praise ("the assassination of martyr commander hajj Imad Moghnieh").

A.4 Coding Examples of Different Sentiment Values

We detail the coding rules that we developed in the Coding Process section. We provide examples of sentences from our corpus and how we code them below to demonstrate the use of these rules. The sentence's actor of interest, the entity towards which we code sentiment, is italicized.

Negative Sentences (-1)

1. He also shed light on the vicious attitude of the groups fighting in Syria, how the ISIL and *Qaeda* are executing socalled Free Syrian army commanders and even chopping off their heads, asking if such is a demonstration to the values of a revolution.
2. Moreover, *Qaeda* also killed a great number of Sunnite and Shiite scholars.
3. The families of our martyrs are proud of their martyrs because they know the truth of the battle in which these martyrs are fighting, and they do not listen to all those who want us to sit in our homes so that Daesh, *Qaeda*, and the Takfiris, the Americans and the Saud Dynasty behind them, would come to destroy the entire region, crush its souls and stones as they are doing in Yemen.
4. This grave desecration comes just days after the Washington summit, which allowed the enemy Prime Minister *Benjamin Netanyahu* a criminal who leads the usurper entity which abuses and rapes the rights of the Palestinian people under American backing to wear the mask of someone striving for peace, after the Obama administration bowed to the whims of the Zionist expansionist settlement policy.

Neutral Sentences (0)

1. I myself do not really know whether *Netanyahu* really said that or not.

2. Moreover, Palestinian Authority President Mahmoud Abbas, who reached an agreement with *Hamas* in April that led to the formation of a unity government last month in June, welcomed the proposal and urged its acceptance, WAFA said.

3. Until a year and a half ago its pledge of allegiance was to *Qaeda*.

Positive Sentences (1)

1. Our relation with *Amal* became a model in Lebanon and the region.

2. He and the *Amal* is our partner in the battle, in politics, in negotiations, in fighting, in the battlefield, in pain, in the suffering of wounds, and in sufferings.

3. On this level, Sayyed Nasrallah confirmed that the relationship between *Amal* and Hizbullah is based on trust, mutual respect, trust, dialogue and almost daily talk.

4. It is a day for our Jihadi and heroic leader *Haj Imad Moghniyeh*, a day of all leader martyrs who filled different posts in this Resistance and held multiple responsibilities in the different regions, sectors, combat units who are an honor to us.

Sentiment Values

Cross-Validation Data (n=500)		
Negative (-1)	**Neutral (0)**	**Positive (1)**
151	117	232

Table A.1: Distribution of sentences by sentiment in the cross-validation set.

B Sentiment Classification

B.1 Model Information

Our sentiment classifier uses logistic regression (Logit). The Logit uses a linear combination of the features to assign probabilities to each class. Specifically, we use a Logit with an $l2$ penalty. There is a single parameter λ the shrinkage for the coefficient values for this Logit. We perform cross-validation on a full grid search of λ values over a range from 1 to 100000, to select the highest performing model. Our classifier is implemented with the Python package `scikit-learn`.

B.2 Precision, Recall, F-1, Confusion Matrix

	Precision			**Recall**			**F1-Score**		
Label	**-1**	**0**	**1**	**-1**	**0**	**1**	**-1**	**0**	**1**
	0.89	0.70	0.76	0.83	0.43	0.94	0.86	0.53	0.84

Table B.1: Evaluation Metrics for Sentiment Classifier

$$\begin{bmatrix} 126 & 8 & 17 \\ 15 & 50 & 52 \\ 1 & 13 & 218 \end{bmatrix}$$

Figure B.1: Confusion Matrix for LG - Main Analysis

C Additional Analyses

C.1 Comparison to Other Sources of Conflict Data

	Coverage	Years	Actions
ACLED	Africa, S. Asia Middle East	1997-2018 2010-2018 2017-2018	violence, non-violence
SCAD	Africa, Latin America	1990-2016 1990-2016	violence, non-violence
UCDP	Global	1989-2016	violence

Table C.1: Coverage from other conflict data sources

Conflict data at the subnational data has increased both in coverage and depth of content over the last several years. In particular the ACLED, SCAD, and UCDP's Non-state Actor data are widely used in the field. However, while they posses many advantages, they are ill-suited for the analysis proposed in this paper. In the following section we detail our attempt to measure Hezbollah's alliances and rivalries using these widely available sources of conflict data. We should be clear here that this is not a critique of the data, only to the application of this particular problem. Issues arise either because the data does not cover the given region (SCAD) or because the data has only recently began coding data in the region (ACLED). This is not the fault of the teams involved. They have put together high quality, continuously updated, and publicly available sources of data. We simply note that there are advantages to using a text based approach for this specific question.

While it could be said that the topic we have chosen is narrow, previous studies find that alliances and rivalries between non-state groups may lead to longer civil conflicts (Horowitz and Potter, 2014; Phillips, 2014) and increased violence against vulnerable civilians (Wood and Kathman, 2015). This suggests a need to further study the subject using any available approach, especially if they can be demonstrated to have some advantages. It is also important to note that Hezbollah is not a obscure group. They have existed for decades, operated in several countries, conducted high profile attacks, and are a major political party in Lebanon. This not only makes this specific case worth studying, but also suggests that conducting an analysis of rivalries and alliance between more obscure actors would be difficult or impossible using the available data.

The UCDP Non-State Conflict (Sundberg et al., 2012) data specifically aims to collect data on the interactions between non-state actors. While it provides global coverage and data from 1989 through 2016, the focus of this data is on violent actions between non-state groups. While the use of violence is often the focus of studies in political science, if the data only codes instances of violence, it is difficult to measure the intermediate stages of escalation of rivalries between actors. Similarly, because the data focuses on violence between actors it can not be used to measure alliances between non-state groups. The table below shows the data available for our actor of interest Hezbollah. While the data does cover Amal for a few years during their conflict with Hezbollah, it does not cover their alliance in the preceding decades (this is also true for the PSP). The relationship between Hezbollah and Hamas has not been violent in nature. They are consequently not included in the UCDP data. The data does capture the violence between al-Qaeda-linked group al-Nusra Front, as does our approach, but we are unable to observe the slow degradation of this relationship over time. Overall, the UCDP data has global coverage and a longer time frame, and is a great source of data for violence between non-state groups, but it does not contain information on non-violent contentious actions such as threats or condemnation and it does not contain data on alliances.

Actor	Years
Amal	1989-1990
SLA	1992-1999
PSP	2008
Brigades of Aisha	2013
Al-Nusra Front	2013-2015
ISIS	2014-2015

Table C.2: Actors and Years Covered in UCDP Non-state Actor Data

We now consider the Social Conflict Analysis Database (SCAD) (Salehyan et al., 2012). As noted in their codebook, SCAD "contains information on protests, riots, strikes, and other social disturbances in Africa, Latin American and the Caribbean. Whereas conflict data is generally available for large-scale events such as civil and international war, the purpose of this dataset is to compile information on other types of social and political disorder."[6] SCAD codes an impressive number of both violent and non-violent actions, but currently only covers Africa, Latin America, and the Caribbean. This makes it unusable for the actors involved in this study. While it would be possible to conduct a similar analysis using a set of African groups, the analysis in the main paper using ICEWS demonstrates that there are advantages to using the text of non-state actors.

The Armed Conflict Location and Event Data Project (ACLED) is a similar conflict data collection project (Raleigh et al., 2010). They record the source, time, type, location, and effect of both violent and non-violent political events. They focus on Africa, Asia, and the Middle East. Similar to SCAD, ACLED codes a number of different actions that both state and non-state actors engage in. The data expanded coverage from Africa and South Asia to also include the Middle East in 2017. While it is impressive that the data is updated in near real time, we are unable to conduct the analysis in our article using ACLED data

[6]SCAD Codebook

currently. 2017 and parts of 2018 are the only dates available for Hezbollah activities. As ACLED continues its collection process it is likely that in the future it may serve as a valuable source of data that could likely be supplemented with the text of non-state groups to produce a more comprehensive measure of the interactions between non-state actors.

C.2 ICEWS Data - Verbal Events

We use the Integrated Crisis Early Warning System (ICEWS) event dataset[7] to model relationships between non-state actors and compare the results to our main analysis' findings (O'Brien, 2010; Boschee et al., 2017). ICEWS is a DARPA-funded initiative that uses hundreds of newspaper sources to automatically collect and code events data. The ICEWS dataset covers the years 1999-2017. We select events where Hezbollah is the source (the actor carrying out the event), while either al-Qaeda, Amal, or Hamas are the target of the event. The names of the groups are standardized to ensure all mentions are captured across time. We match the CAMEO code of each event to its QuadClass[8] so that each event is coded as either neutral, verbal cooperation, material cooperation, verbal conflict, or material conflict. We then create a measure that contains only the verbal events from ICEWS, which matches our data more closely, and another which includes all the events. Next, the QuadClass is collapsed into three categories (positive, neutral, and negative). This allows a more direct comparison to our main analysis. Consequently, Hezbollah praising Amal is positive and coded 1, while Hezbollah threatening al-Qaeda is negative and coded -1. As with our main analysis, we take the yearly average of these scores then apply the Kalman filter to generate smoothed estimates of Hezbollah's sentiment towards each actor.

[7]The ICEWS data and codebook can be found here- https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/28075

[8]We use the example provided by the Open Event Data Alliance- http://phoenixdata.org/description

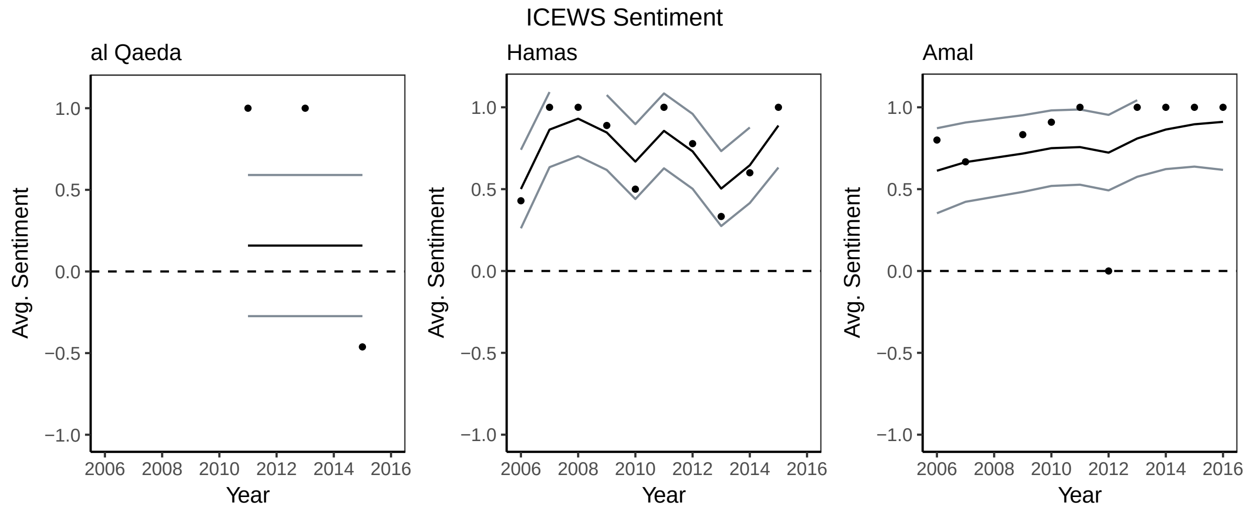

Figure C.1: ICEWS Verbal Events

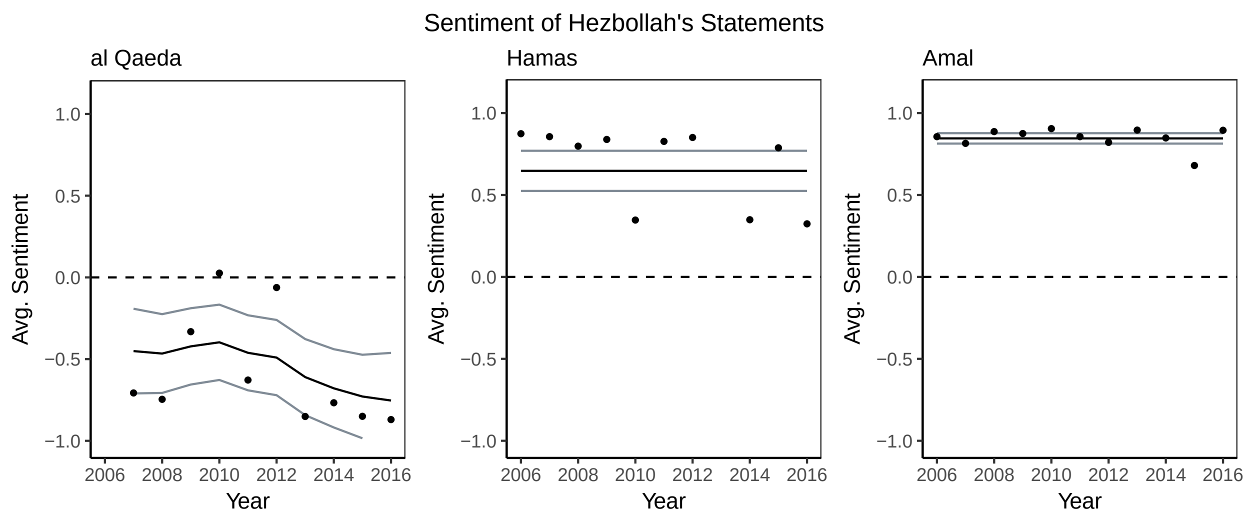

Figure C.2: Main Analysis

We now compare our approach to only the verbal events from ICEWS. As with the analysis in the main paper, we find consistent results for Hezbollah's sentiment towards Hamas and Amal- both are consistently positive. However, there is considerably more variation in the verbal ICEWS events, than our analysis, or the analysis of the full ICEWS data. Again the largest difference between our measure and the one derived from ICEWS in that of al-Qaeda. While the verbal ICEWS events does not show the same shift from perfectly positive to perfectly negative like the analysis using the full ICEWS events, there are important differences. First, there are only three years of recorded data, while a longer time frame (10 years) can be coded from Hezbollah's official documents. Based on the ICEWS verbal measure, it appears that Hezbollah was largely indifferent to al-Qaeda during a period of time in which the two groups are openly on opposite sides of the Syrian Civil War. Additionally, the width of the 90% confidence interval raises the potential that the relationships between Hezbollah and al-Qaeda are statistically indistinguishable from those of an ally, such as Amal or Hamas.

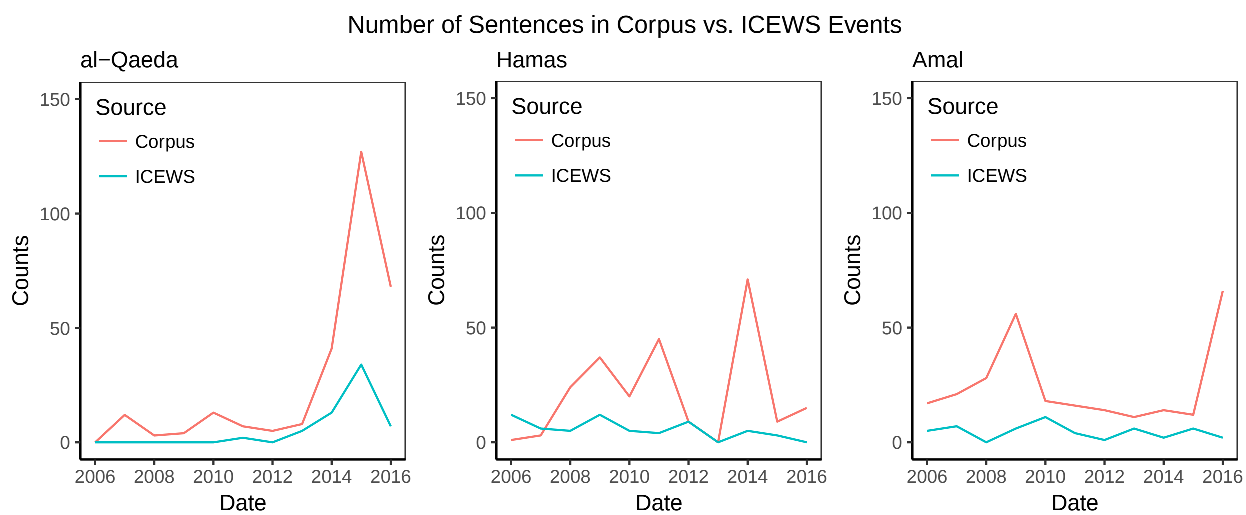

Figure C.3: A comparison of the number of actor sentences in our corpus of official Hezbollah documents vs. events in ICEWS.

C.3 Accounting for Actor Names

The sentences we capture often contain the name of one of our actors of interest, so it is possible that we have built a classifier for these actor names rather than sentiment more broadly. While this would not necessarily impact the results presented in the paper, it might impact the generalization of the classifier. One approach is to remove the names of each of these actors from the corpus before coding and classifying. Doing this decreases the accuracy of the classifier from 80% to 70%. That the sentiment classifier does worse after removing information is unsurprising, but it is worth ensuring that the accuracy is still sufficiently high, even without actor names. In the Model Comparison section we also evaluate this approach using 1000 random draws of the training and test sets, to further ensure our results are not driven strictly by the inclusion of actor names. Taken together, the actor names do have some impact on the accuracy of the model, but our tests show little evidence that they are driving our results.

	Precision			**Recall**			**F1-Score**		
Label	**-1**	**0**	**1**	**-1**	**0**	**1**	**-1**	**0**	**1**
	0.71	0.59	0.71	0.64	0.35	0.90	0.68	0.44	0.79

Table C.3: Evaluation Metrics for Sentiment Classifier Without Actor Names

$$\begin{bmatrix} 97 & 19 & 35 \\ 24 & 41 & 52 \\ 15 & 9 & 208 \end{bmatrix}$$

Figure C.4: Confusion Matrix for LG - No Actor Names

C.4 Further Validation: Moving Beyond Non-State Groups

To further validate our method and demonstrate its use outside of non-state groups, we classify sentiment on two additional political actors ("anchors") that we expect to have consistent sentiment, whether positive or negative, across the *entire* corpus (no variation expected). We use "Imad Mughniyah" as the positive anchor. Mughniyah was one of Hezbollah's first members and a principal actor within the organization until his assassination in 2008 (Levitt, 2008). He is believed to have planned and executed a range of terror attacks, including the infamous Beirut Barracks Bombing and US Embassy Bombing in Beirut. These actions and a host of others led the United States to offer a bounty of as much as $25 million for information leading to his capture (Gleis and Berti, 2012). After his assassination, Hezbollah's leadership quickly and publicly mourned Mughniyah (Stern, 2008). He is referenced in very positive terms in our corpus across all years and lionized as a martyr.[9] This is seen clearly in the following passage from our corpus- "There is nothing left to be accomplished since Hajj Imad [Mughniyah] in his humble spirit, his specialties and brilliant mind, has become a title and a symbol of the very advanced stage of the resistance, in its advanced development and ability to confront assault and aggression."

We use the three Israeli prime ministers (Ariel Sharon, Benjamin Netanyahu and Ehud Olmert) that served during the time range of our corpus as the negative anchor. We aggregate the sentiment towards these individuals for our analysis. One of Hezbollah's core missions is to fight Israel and the group has engaged in open warfare with the state (Avon et al., 2012). Since the Prime Minister is the leader of the State of Israel, they are a natural

[9] Hezbollah released a statement, which is included in our corpus, a day after Mughniyah's death saying: "After a life full of Jihad, sacrifices and accomplishments lived with a longing to martyrdom, Islamic Resistance leader Hajj Imad Moghnieh was assassinated by Israeli criminal hands."

target for negative rhetoric and criticism related to the state's actions. As such, we expect sentiment towards Israeli prime ministers to be consistently negative with no variation. This passage from our corpus exemplifies the negative sentiment directed towards these individuals- "This grave desecration comes just days after the Washington summit, which allowed the enemy Prime Minister Benjamin Netanyahu a criminal who leads the usurper entity which abuses and rapes the rights of the Palestinian people under American backing to wear the mask of someone striving for peace..."

We perform the same analysis detailed in the paper[10] and display our results in figure C.5. Hezbollah's relationship with each of the actors in our analysis is consistent with their average sentiment trends. The sentiment towards "Imad Mughniyah" across all group years is positive. Similarly, the sentiment of Hezbollah's speech about Israeli Prime Ministers is consistently negative. This provides additional evidence that our sentiment classifier can correctly classify the sentiment of speech about political actors. It also demonstrates the use of this method to perform analyses beyond modeling relationships between non-state groups.

[10]We code the sentiment of 300 sentences containing one of the standardized names of these anchors to train the classifier for this analysis.

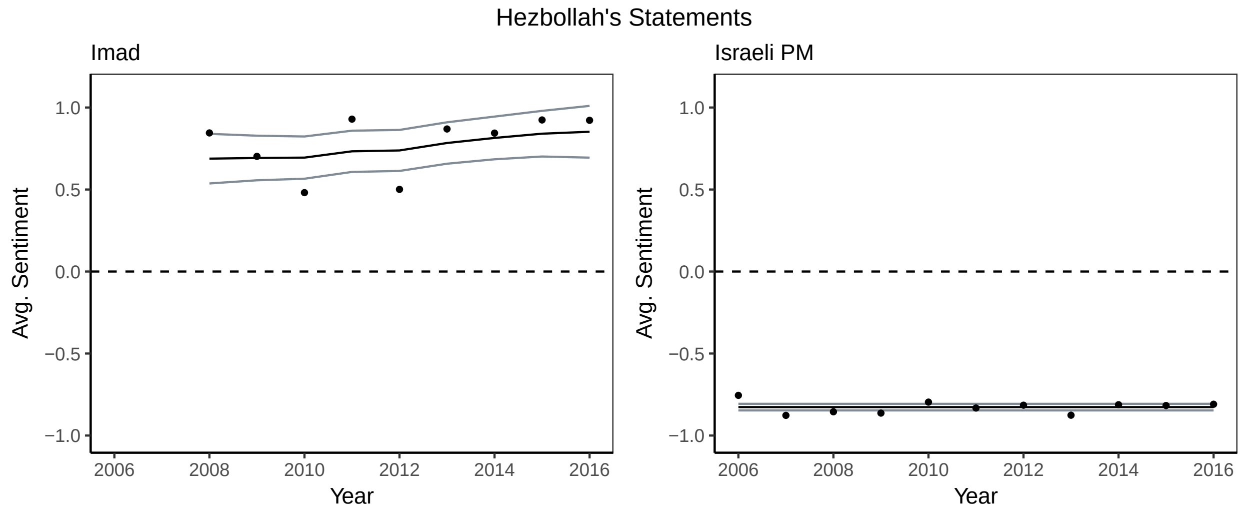

Figure C.5: The points represent observed sentiment, while the black lines represent smoothed Kalman estimates. The 90 percent confidence interval is marked by the gray lines.

C.5 Model Comparisons

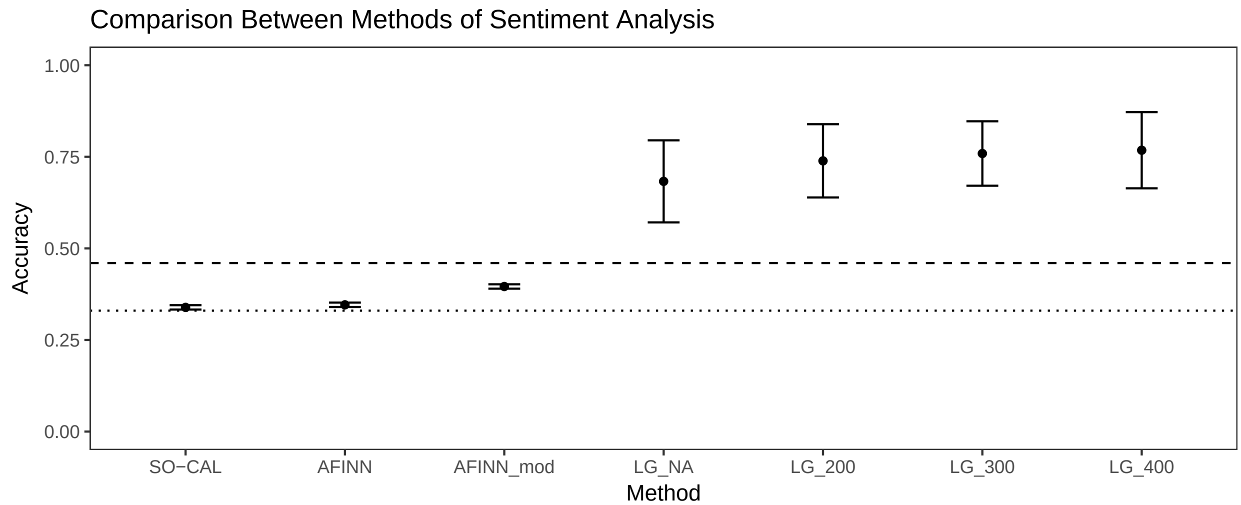

Figure C.6: Comparison of accuracies across various methods of classification. LG_400 (400 sentences in training set) is used in the main analysis. The doted line marks the accuracy for random guessing (33 percent), while the dashed line marks the accuracy for majority class guessing (46 percent). The machine learning approach consistently produces better results than a dictionary approach, and the accuracy remains for smaller training sets, and when actor names are removed.

In this section we compare the accuracy of our sentiment classifier to common dictionary-based methods of classifying sentiment, finding our approach classifies sentiment with considerably higher accuracy. We also find that our choice of training set size has no significant impact on the average accuracy of our classifier. This suggests not only that the model is robust to training size, but that the task could be done effectively at lower training sizes. Finally we test the accuracy of the model after removing actors names (LG_NA).

A common approach to sentiment analysis uses pre-made dictionaries that contain a positive or negative sentiment for thousands of words. Methods that use this procedure have been found to do well in classifying text such as movie reviews. However, these dictionaries are often context-specific and usually perform poorly outside of their domain

(Grimmer and Stewart, 2013). To our knowledge, no dictionary has been created to classify the text of violent political actors. In order to compare our classifier with this approach, we use two popular sentiment dictionaries: AFINN-111 (Nielsen, 2011) and the SO-CAL dictionary (Semantic Orientation Calculator) (Taboada et al., 2011). The AFINN-111 contains thousands of words and their positivity or negativity. The SO-CAL dictionary contains over 10,000 entries and accounts for lexical features such as negation. As a final test, we add positive and negative words that are specific to our corpus (takfiri, terrorist, etc.) to the AFINN-111 dictionary.

The dictionary methods sum the sentiment scores (1,-1) of all the words in a sentence to measure its sentiment. If the sum is greater than 0, the sentence is classified as positive. If it is less than 0, it is classified as negative. The classifier method, which utilizes a Logit), uses the term frequency–inverse document frequency (tf-idf) counts to learn the mappings between the word features and the sentiment scores. To compare between the dictionary methods and our classifier, we draw 100 random sentences with replacement and then calculate their sentiment using each dictionary method. This process is repeated 1000 times to create a measure of the average accuracy of each approach as well as a confidence interval. Figure C.6 shows that our machine learning approach (LG) does substantially better than any of the dictionary-based approaches. The horizontal dashed line represents the accuracy of random guessing for a three class problem (33%). Both of the non-modified dictionaries have an accuracy near that of random guessing.

As an additional test of the effectiveness of our model we first use the model trained on sentences without the actors names included (LG_NA) and calculate the same measure of average accuracy and confidence intervals as the other Logit models. While there is some decrease in accuracy this approach is still considerably better than the dictionary methods, or majority guessing. After running analyses on data that removes actor names, it is safe to say that the results of our classifier are not driven solely by the presence of actor names.

Further, we also compare the accuracy of our Logit model across different training sizes. In these analyses we again draw 100 random sentences to be used as the test set, but here we also randomly draw either 200, 300, or 400 sentences to be used for the training set. This process is done 1,000 times for each training set size to estimate the mean accuracy and confidence interval. This provides an additional estimation of the generalization error from the 10-fold cross validations presented earlier. The confidence intervals around the three Logit models shows that there is not a significant difference in the accuracy of the classifier across different training sizes. Since classifying sentiment towards political actors is a difficult task, we employ the model with 400 sentences in our main analysis, but there is little evidence that using a larger or smaller training set would significantly alter our results.

D References

Alagha, J. (2006). *The Shifts in Hizbullah's Ideology*. Amsterdam University Press.

Avon, D., A.-T. Khatchadourian, and J. M. Todd (2012). *Hezbollah: A History of the "Party of God"*. Harvard University Press.

Boschee, E., J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, and M. Ward (2017). Icews coded event data.

Gleis, J. and B. Berti (2012). *Hezbollah and Hamas: A Comparative Study*. Johns Hopkins University Press.

Grimmer, J. and B. M. Stewart (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis 21*(3), 267.

Horowitz, M. and P. Potter (2014). Allying to kill terrorist intergroup cooperation and the consequences for lethality. *Journal of Conflict Resolution 58*(2), 199–225.

Khatib, L., D. Matar, and A. Alshaer (2014). *The Hizbullah Phenomenon Politics and Communication*. Oxford University Press.

Lee, H., Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu, and D. Jurafsky (2011a). Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Conference on Natural Language Learning (CoNLL) Shared Task*.

Lee, H., Y. Peirsman, A. Chang, N. Chambers, M. Surdeanu, and D. Jurafsky (2011b). Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pp. 28–34. Association for Computational Linguistics.

Levitt, M. (2008). *Hamas: politics, charity, and terrorism in the service of jihad*. Yale University Press.

Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. New York: Cambridge University Press.

Manning, C. D., M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60.

Nielsen, F. Å. (2011, mar). Afinn.

O'Brien, S. P. (2010). Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review 12*(1), 87–104.

Phillips, B. (2014). Terrorist group cooperation and longevity. *International Studies Quarterly 58*(2), 336–347.

Raghunathan, K., H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky, and C. Manning (2010). A multi-pass sieve for coreference resolution. In *Empirical Methods in Natural Language Processing (EMNLP)*.

Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing acled: an armed conflict location and event dataset: special data feature. *Journal of peace research 47*(5), 651–660.

Recasens, M., M.-C. de Marneffe, and C. Potts (2013). The life and death of discourse entities: Identifying singleton mentions. In *North American Association for Computational Linguistics (NAACL)*.

Salehyan, I., C. S. Hendrix, J. Hamner, C. Case, C. Linebarger, E. Stull, and J. Williams (2012). Social conflict in africa: A new database. *International Interactions 38*(4), 503–511.

Soon, W. M., H. T. Ng, and D. C. Y. Lim (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics 27*(4), 521–544.

Stern, Y. (2008, February). Nasrallah: Mughniyah's blood will lead to israel's downfall. *Haaretz*.

Sundberg, R., K. Eck, and J. Kreutz (2012). Introducing the ucdp non-state conflict dataset. *Journal of Peace Research 49*(2), 351–362.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll, and M. Stede (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics 37*(2), 267–307.

Wood, R. and J. Kathman (2015). Competing for the crown. *Political Research Quarterly 68*(1), 167–179.