

INTRODUCTION

- What is a dictionary?** Topic-based dictionaries are an important linguistic tool to help researchers detect the prevalence of latent constructs. We devised a dictionary that diagnoses how often threat is evoked in texts using word2vec’s trained model, with several validation methods.
- Why a threat dictionary?** Rises in collective threat (e.g., war, natural disasters, and pathogens) have been linked to important shifts in societal norms such as the tightening of cultures. Tight cultures have strong norms and less tolerance for deviance and are linked to lower creativity and openness but higher order.¹

H1: Linguistic indicators of threat will decrease over time.²
H2: Changes in linguistic indicators of threat will relate to greater occurrences of words related to cultural tightness, natural disaster events, and number of homicides.
H3: The threat dictionary will be able to indicate between-group differences among Republicans and Democrats.

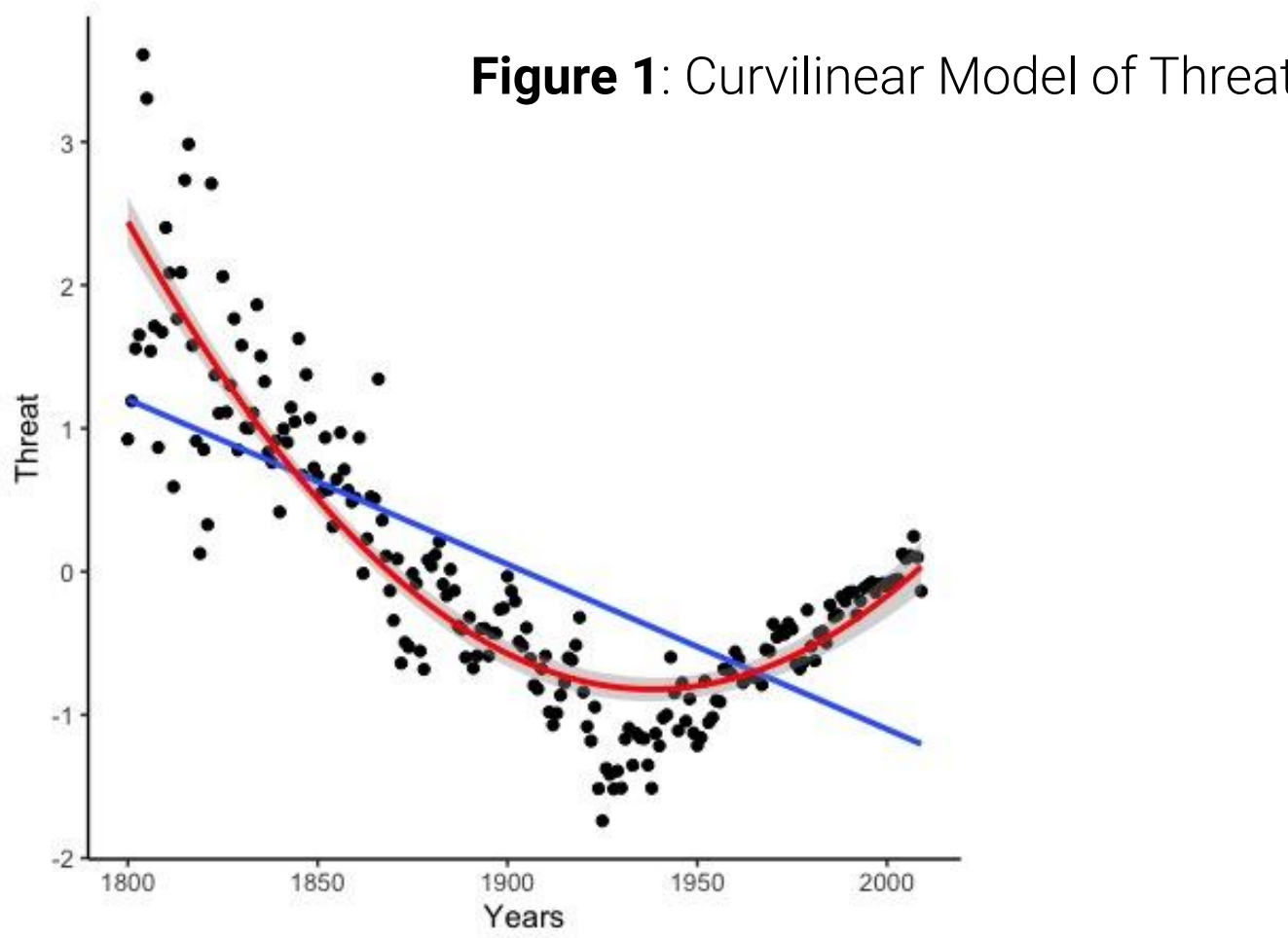
METHODS

1. **Creating the Threat Dictionary.** Based on Gelfand et al. (2011)’s tight-loose theory (TL), we gathered keywords for the threat categories of natural disasters, pathogens, and human-made conflicts. Using word2vec, we expanded this list based on proximal coordinates in a 300-dimension mapping of the GloVe corpus.³ The 18-word index of our threat dictionary (See Table 1 for examples) were derived from the resulting clusters that were sorted, then selected based on sharedness across the key threat categories.

2. **Validation Data Sources.** Google Books provides a percentage of words found per year within its digital archive of books published from 1600 to 2008. We used this source to collect the relative frequency of threat words from the last 200 years. The same data collection process was used for our tightness variables with words provided by the tight-loose dictionary.⁴ For additional validation purposes, our other sources of data were NOAA’s number of natural disaster events,⁵ the UN’s International Homicide Statistics,⁶ threat words per year in Newspapers.com,⁷ 400+ speeches from U.S. presidents in the post-war era, and 76 2016 campaign speeches made by presidential candidates Hillary Clinton and Donald Trump.

RESULTS

Is threat going down? Our linear model showed an overall decrease in threat words: ($R^2 = 0.49$). However, a quadratic model proved to be a better fit ($R^2 = 0.81$) with this decline changing in recent years, wherein threat words began increasing (See Figure 1). The quadratic model explains an additional 31 percent of the variance. See Figure 2 for threat projections in the next 20 years.



What is threat related to? We examined several indicators as sources of convergent validity. Results of Kendall correlations found (a) the relative frequency of threat and TL words in Google Books were related ($\tau = 0.43, P < 0.001^{***}$); (b) threat numbers in Google Books positively correlated with annual occurrences of natural disaster events from 1980 to 2009, $\tau = 0.33, P < 0.05^*$; and (c) we collected a threat frequency count from an alternative corpus Newspapers.com to demonstrate threat levels and its association with related phenomena, i.e., annual incidents of homicides in the U.S. from 1995 to 2016, $\tau = 0.56, P < 0.001^{***}$. See Figure 3. Additionally, by using Google Books data on words for threat and TL, the results of our regression model ($R^2 = .34, t(1, 208) = 10.36, p < .001^{***}$) found that threat significantly predicted relative levels of tightness ($\beta = 1.06, p < .001^{***}$). This relationship held with lagged models of threat for three different time periods: one year ($R^2 = .34, t(1, 207) = 10.31, p < .001^{***}$), three years ($R^2 = .35, t(1, 205) = 10.56, p < .001^{***}$), and five years ($R^2 = .36, t(1, 203) = 10.64, p < .001^{***}$).

Who uses more threatening language, Republicans or Democrats? U.S. presidential speeches were collected from the post-war period, beginning with speeches from Truman (1945) to Obama (2016). (See Figure 4 for spikes in threat as they relate to major historic events.) These texts were divided into speeches made by Republicans ($n = 164$) and Democrats ($n = 244$). We compared the use of threat words by Republicans ($M = 0.40, SD = 0.31$) versus Democrats ($M = 0.34, SD = 0.28$). The results of our Welch’s Two Sample t-test was $t(330) = -1.99, p < 0.05^*$. Additionally, we contrasted the 2016 campaign speeches of presidential candidates Donald Trump ($n = 51$) and Hillary Clinton ($n = 25$). We compared the percentage of threat words found in Donald Trump’s speeches ($M = 0.33, SD = 0.27$) versus Hillary Clinton’s speeches ($M = 0.18, SD = 0.11$) and found a significant difference between the two contenders, $t(72) = -3.29, p < .01^{**}$. See Figure 5.

DISCUSSION

This dictionary is a tool that provides researchers with important data on threat trends evoked in texts--from newspapers to tweets--across time. With this linguistic measure, researchers and policy-makers can target growing instability in regions around the globe and locate future shifts in socio-political attitudes across societies such as how threat impacts norms, governing preferences, political affiliations, and ethnocentric attitudes. Future validation checks of the threat dictionary will entail replicating the steps used to derive our index of words in another corpus, using Newspapers.com’s extended collection of U.S. newspapers to evaluate additional threat trends in history, and applying the dictionary in different languages and cultural contexts.



Linguistic indicators of **threat** have been in decline over the last **two centuries**, until recently. Threat rates **predict increases** in the strength of **social norms**.

We combined **natural language processing** with **tightness-looseness theory** to create a linguistic **tool** for measuring **changes in threat**.

“

*Our very freedom came under **attack**...
The **victims** were in the airplanes...*

- President George W. Bush, 9/11/2001

”

“

*Ours will be the generation that finally wins the **war** against cancer and begins a revolution in our fight against all **deadly** diseases.*

- President Bill Clinton, 01/27/1998

”

*Sample texts from U.S. presidential speeches indicating use of threat words

Table 1: Sample of Threat Dictionary Words

- attack
 - warning
 - victim
- danger
 - kill
 - serious

Figure 2: Threat Projections for the Next 20 Years

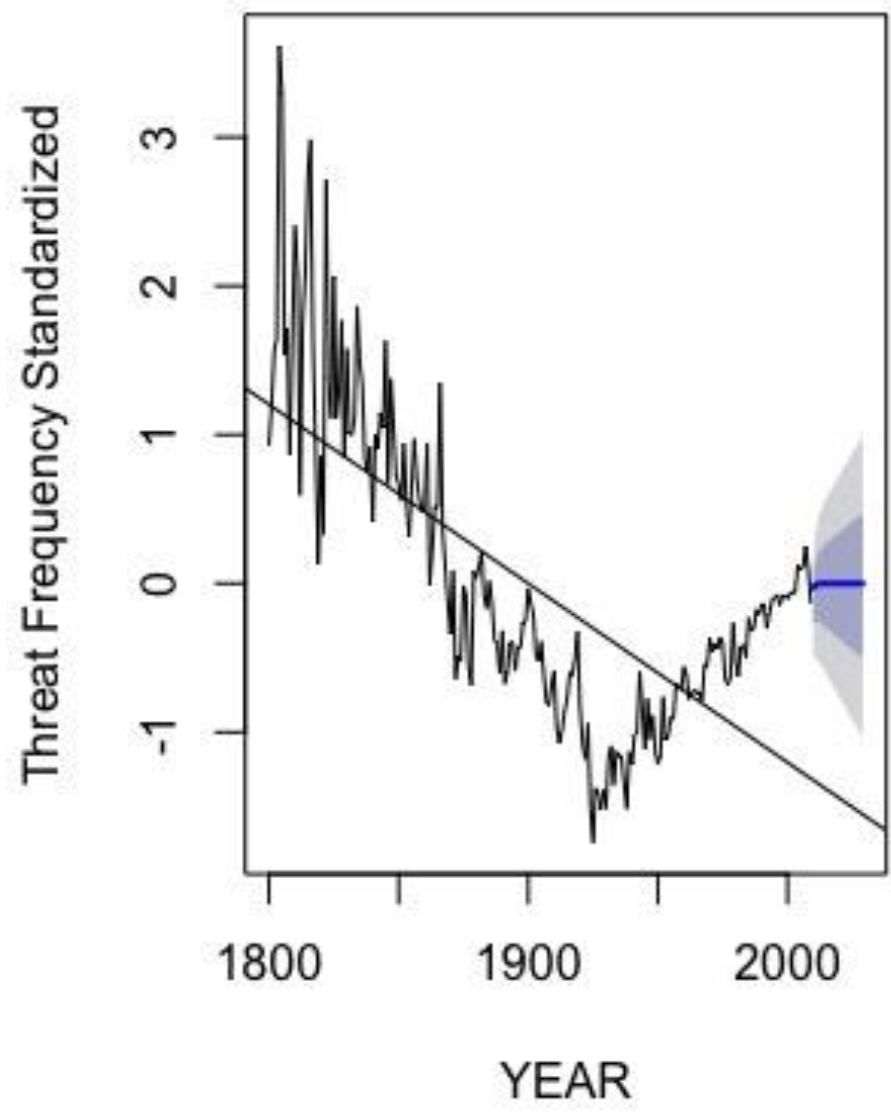


Figure 3: Convergent Indicators

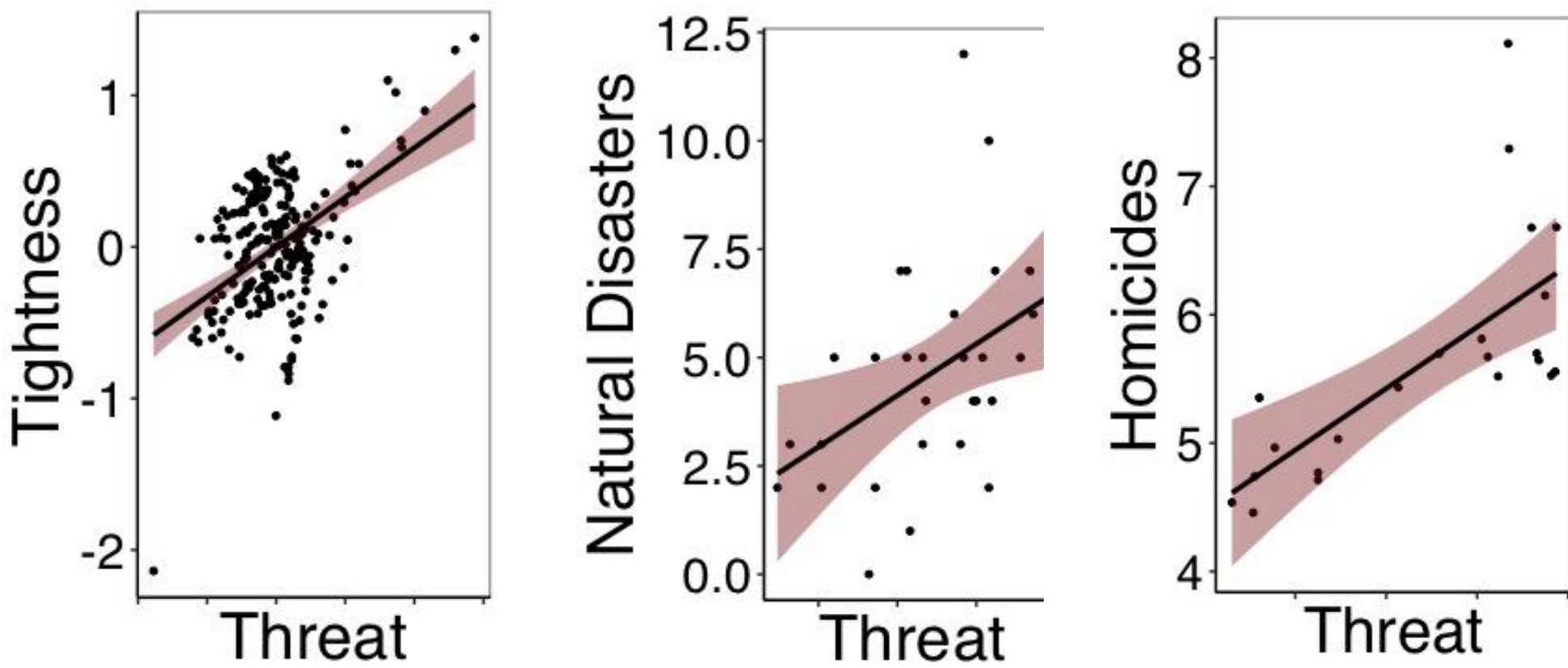


Figure 4: Threat Spikes in U.S. History based on Presidential Speeches

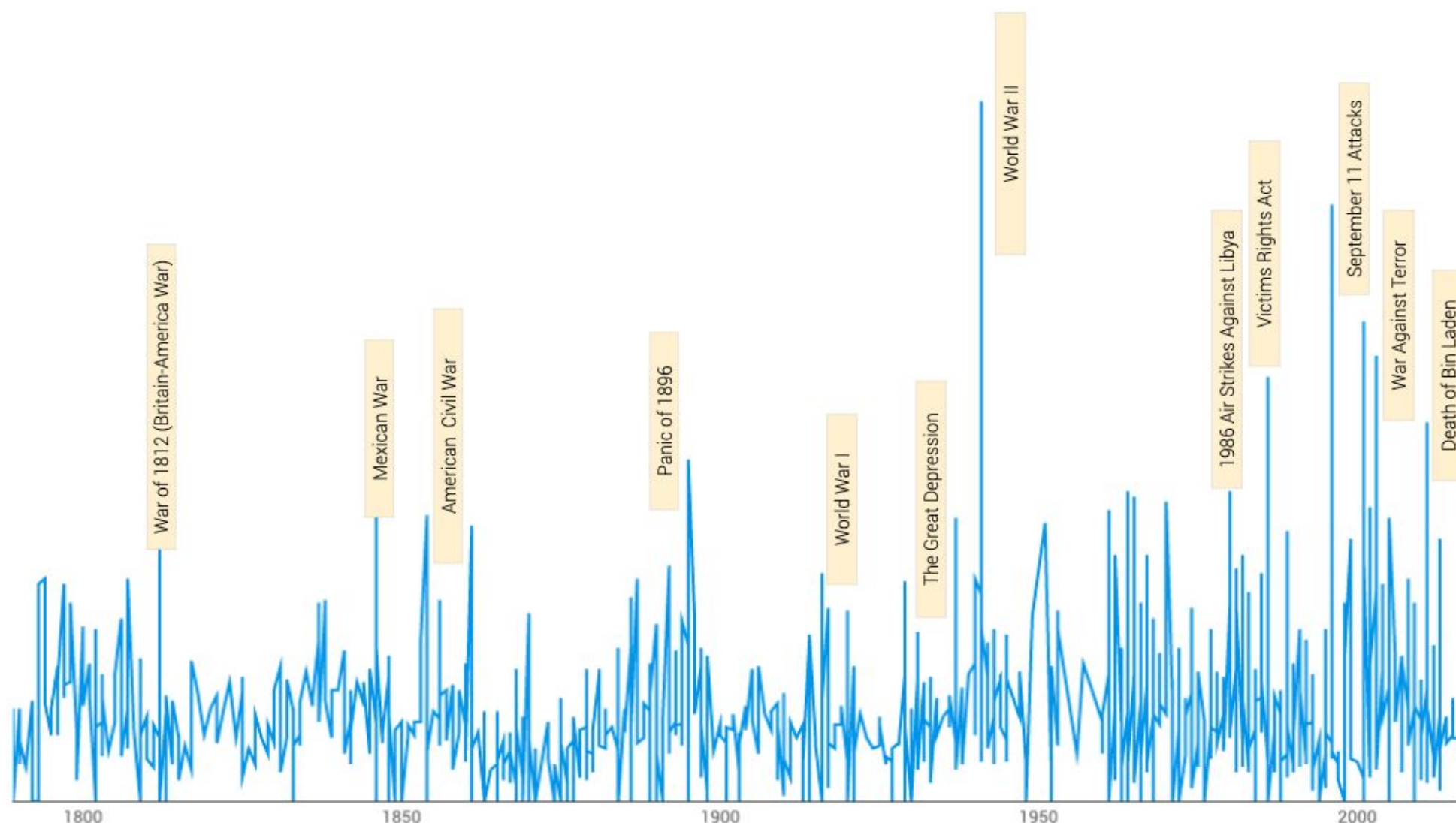


Figure 5: Comparison of Political Speeches

