gov_51_final_project

Owen Bernstein and Lindsey Greenhill

11/15/2020

Project Introduction

The purpose of this project is to examine what factors influence news coverage. We became interested in this general topic after hearing and reading about increasing media polarization between conservatives and liberals. Per work from the Pew Research Center, Americans are increasingly watching news channels that are seen as in line with their own political views, and avoiding those that are not. The result is a media industry that is, for the most part, divided by political ideology. With this in mind, we wanted to see if news channels that are seen as leaning left or right present meaningfully different news coverage.

Project Design

In order to assess potential differences in news coverage amongst channels, we used a data set created by the Internet Archive's Third Eye Project. This data set contains news chyrons - the scrolling captions at the bottom of broadcast images - for four different news channels (MSNBC, CNN, Fox News, and BBC). Although it would be ideal to work with full transcripts of news broadcasts, we did not have access to that data, and we believe chryons are reasonably representative of each channel's coverage. We only included data from MSNBC, CNN, and Fox News, because the data for BBC was problematic in its transcription. The data includes coverage from October 17, 2020 to November 14, 2020.

We performed textual analysis to classify the language each channel used into five categories: Populist, Environmental, Progressive, Conservative, and Immigration Related. We classified the language using pre-existing dictionaries taken from a 2017 study of partisan language in Belgium. It would have been better to use dictionaries created for American politics, but we still think that the ones in the study are relevant and therefore useful. See the word basket section for a more details discussion on the dictionaries.

We ran a linear regression to see if there was a statistically significant relationship between channel and language usage. We also included binary variables of whether or not the coverage was before the election and whether or not the coverage was in prime time in the regression. See the regression section for a discussion on why we thought these variables are important to include.

Initial Hypothesis

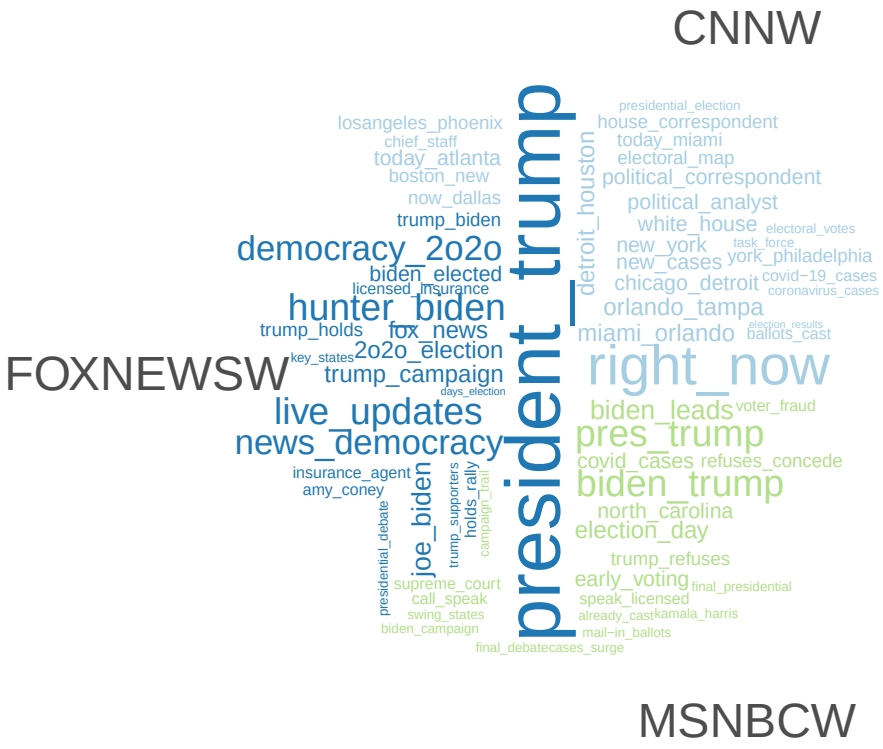
Our initial hypothesis is that Fox News, the most conservative of the news channels, will use more conservative, populist, and immigration related language while CNN and MSNBC will use more progressive and environmental related language.

Exploratory Analysis

We used the quanteda package to perform initial textual analysis of the news chyrons. We also looked at the distributions of language usage across channels. Overall, our findings suggest that our initial hypothesis is incorrect.

Word Cloud Analysis

The graphic below shows a word cloud graphic that compares the language usage between Fox News, CNN, and MSNBC. In our analysis, we excluded filler words, punctuation, etc. and looked for two word phrases. Fox News is dark blue, CNN is light blue, and MSNBC is green.



Discussion

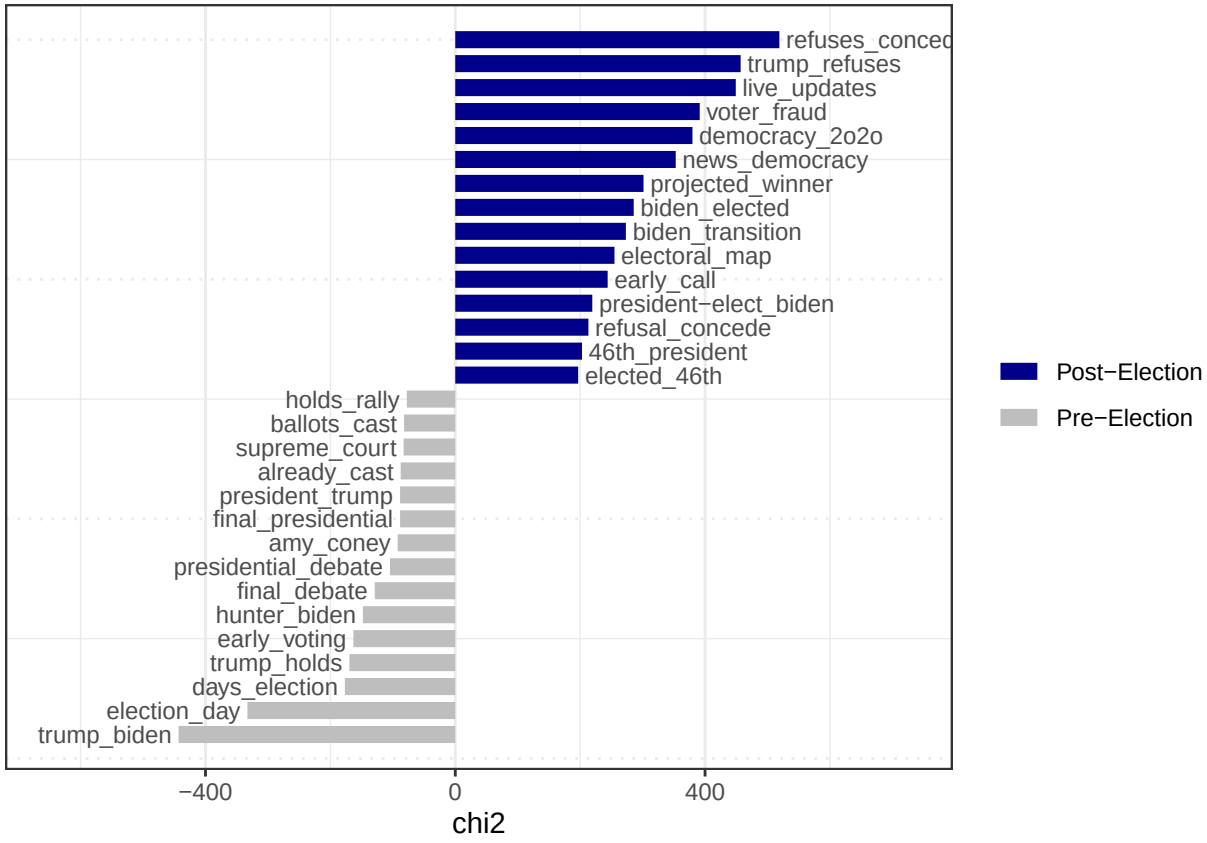
The most mentioned phrase amonst all the networks is President Trump (from Fox), which makes sense given how close our time frame is to the 2020 election. Fox also seemed to focus on Hunter Biden, Joe Biden, and democracy. CNN's most used phrase was "right now," which is perhaps an indication of the tone of their news coverage. CNN also mentions cities such as Chicago, Detroit, Houston, Miami, and a few more. MSNBC also mentions Trump often. It also used the phrases ""biden leads," "north carolina," "election day," and "covid cases" frequently.

Keyness Plot Analysis

In addition to word cloud analysis, we also looked at keyness plots of the news chyrons. A keyness plot is a plot that compares the usage of words between two different data sets. We wanted to look at the potential impact of our post_election and prime_time variables. We created a keyness plot for each variable. The first keyness plot looks at the difference in language use from all channels betwee pre and post election coverage. The second keness plot looks at the difference in language use from all channels between prime time and non prime time coverage.

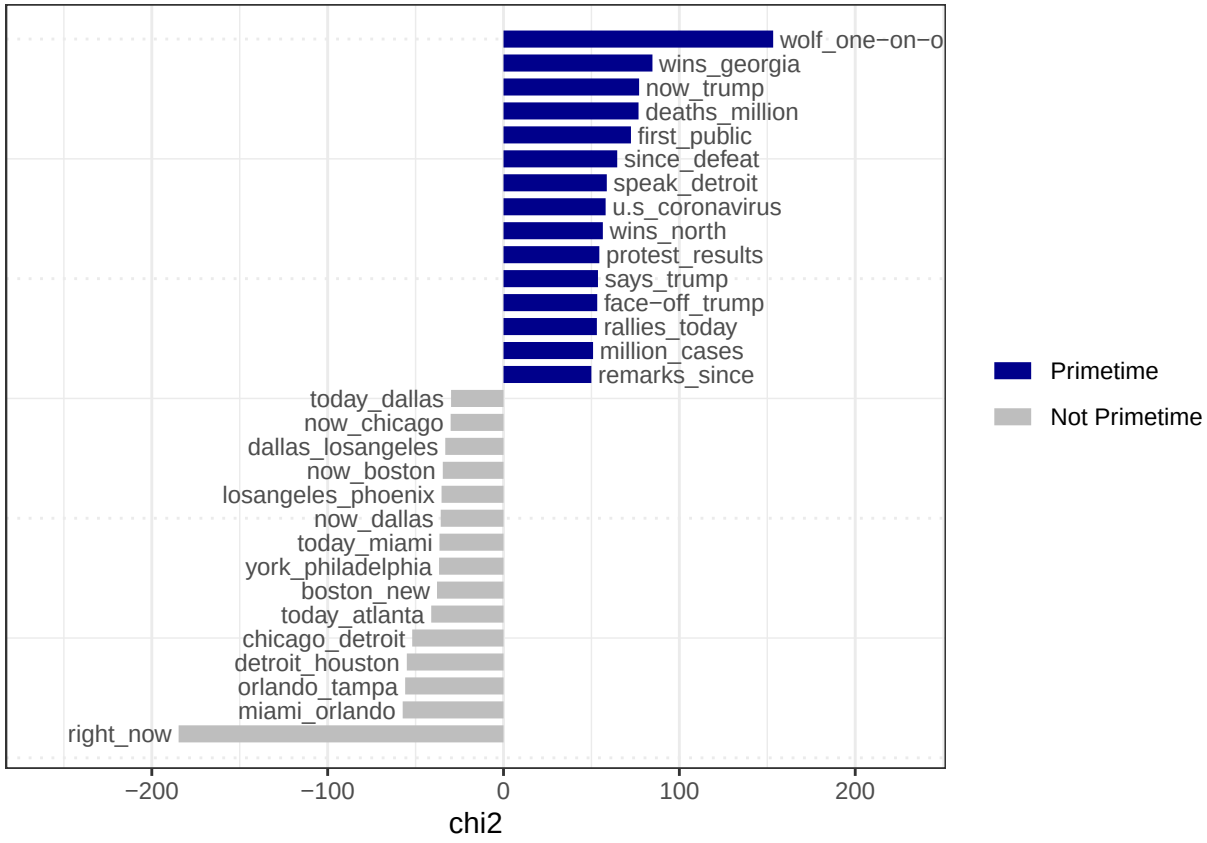
Election Keyness Plot

The differences in language usage pre and post election are intuitive. Compared to after the election, news coverage focused much more on campaigns and voting processes. Compared to before the election, news coverage focused more on calling the election for Biden and the white house transition from Trump to Biden. Overall, there seems to be a shift in topics between pre election and post election coverage.



Prime Time Keyness Plot

We thought it could be interesting to look at prime time coverage vs non prime time coverage because of the saliency of prime time coverage. Additionally, the respective hosts for each channel's prime time news shows tend to be particulary divisive politically, so we thought that might have an impact on the language used during their shows. Interestingly, non prime time coverage seems to focus more on cities compared to prime time coverage. Meanwhile, prime time coverage seems to focus on the election compared to non primt etime coverage.



Boxplot Analysis

Preliminary Analysis

In this analysis, we gathered, cleaned, and wrangled news data from October 17 to November 14, 2020. We chose to include Fox News, CNN, and MSNB in our analysis and assigned those channels ideological rankings of 1, 2, and 3, respectively. The plots below show our preliminary findings.

How much specific language is each channel using?

Visualizing the language each channel uses by a variety of variables.

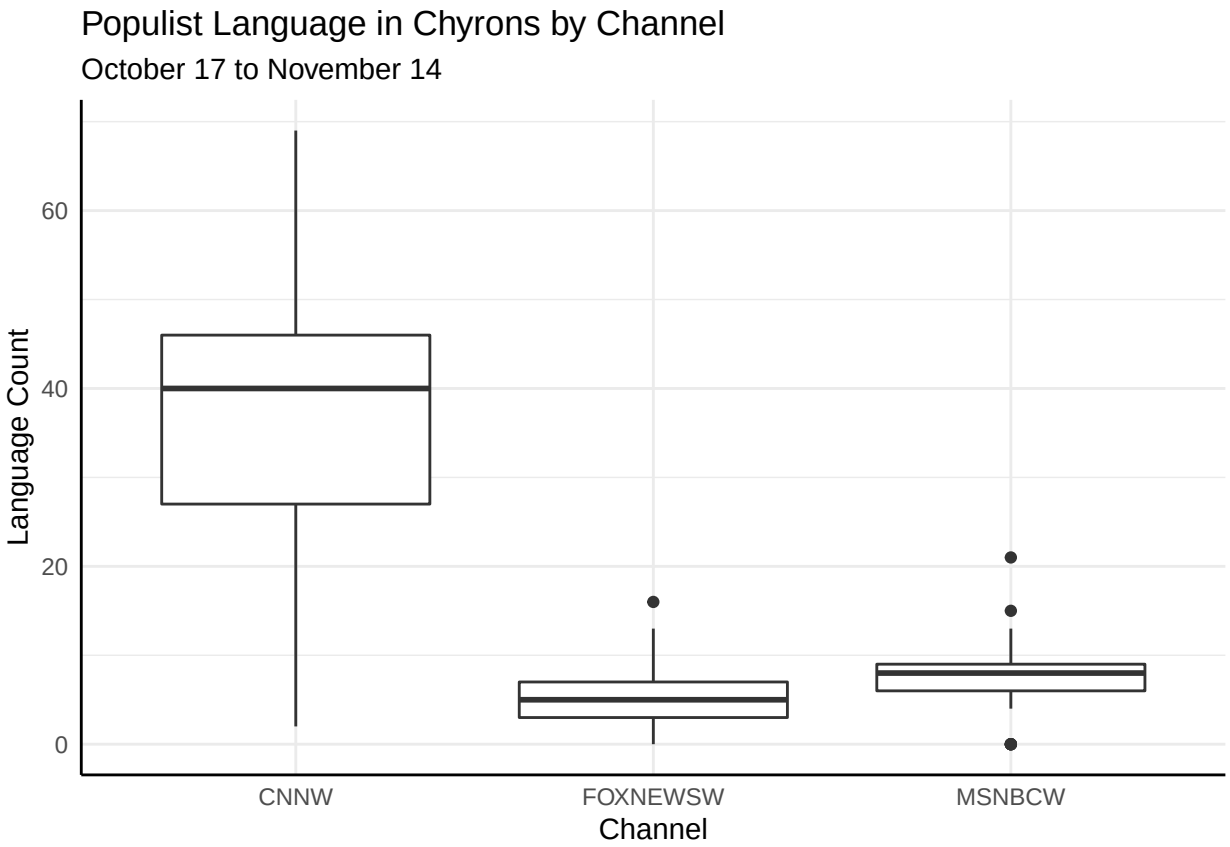
We classified the below categories using word baskets built by the study referenced in our project proposal. The table below shows the raw counts of how many times each channel used a word associated with these categories.

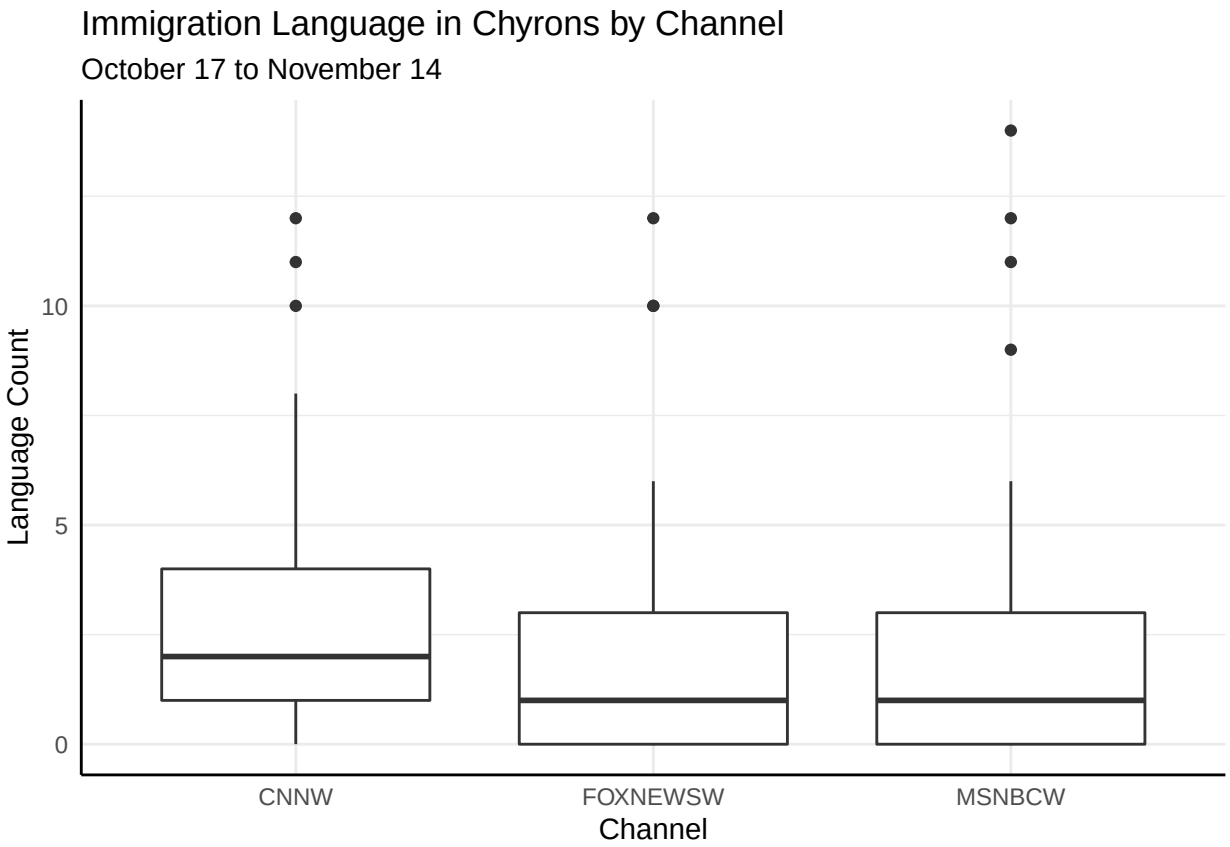
Language Sentiment Across Channels
October 17 to November 14

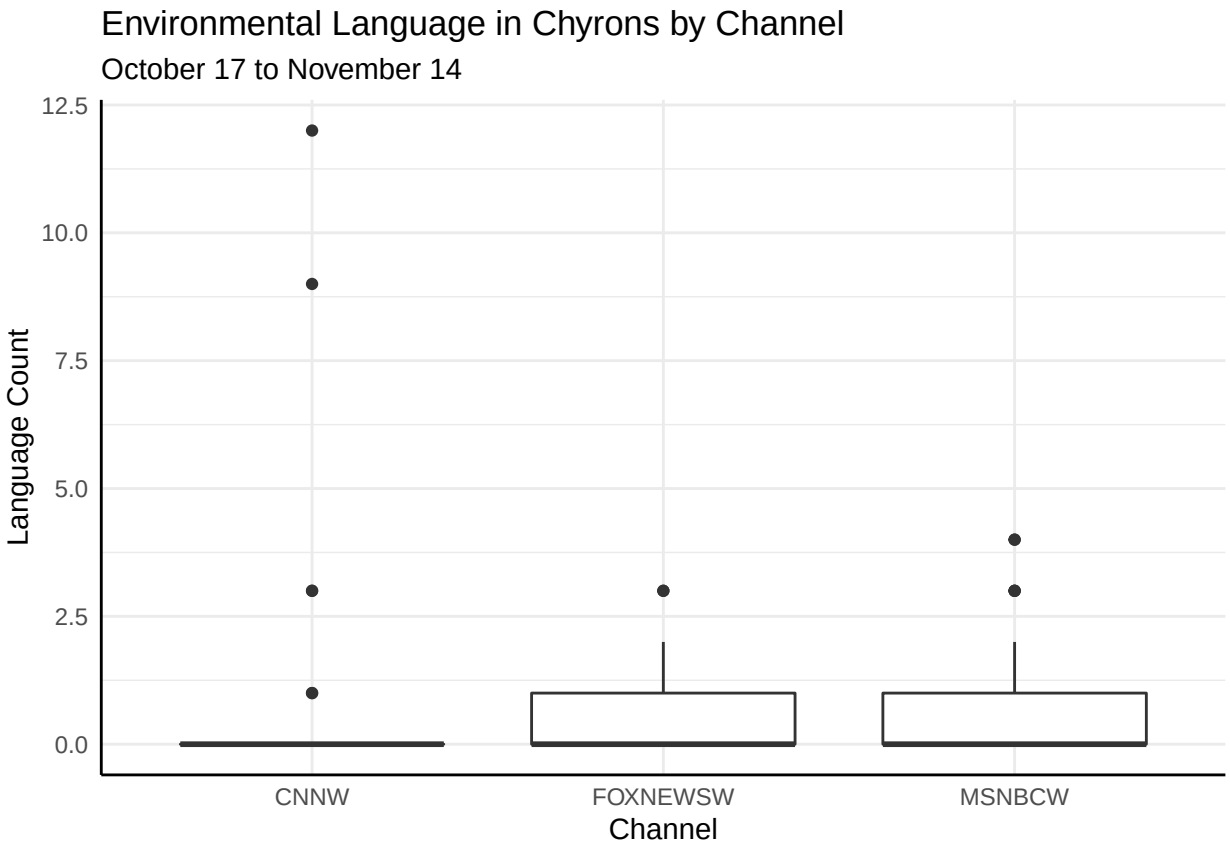
	Word Count for Each Content Category				
Channel	Populism	Environment	Immigration	Progressive	Conservative
CNNW	1022	29	98	940	35
FOXNEWSW	163	15	74	107	10
MSNBCW	219	24	77	80	35

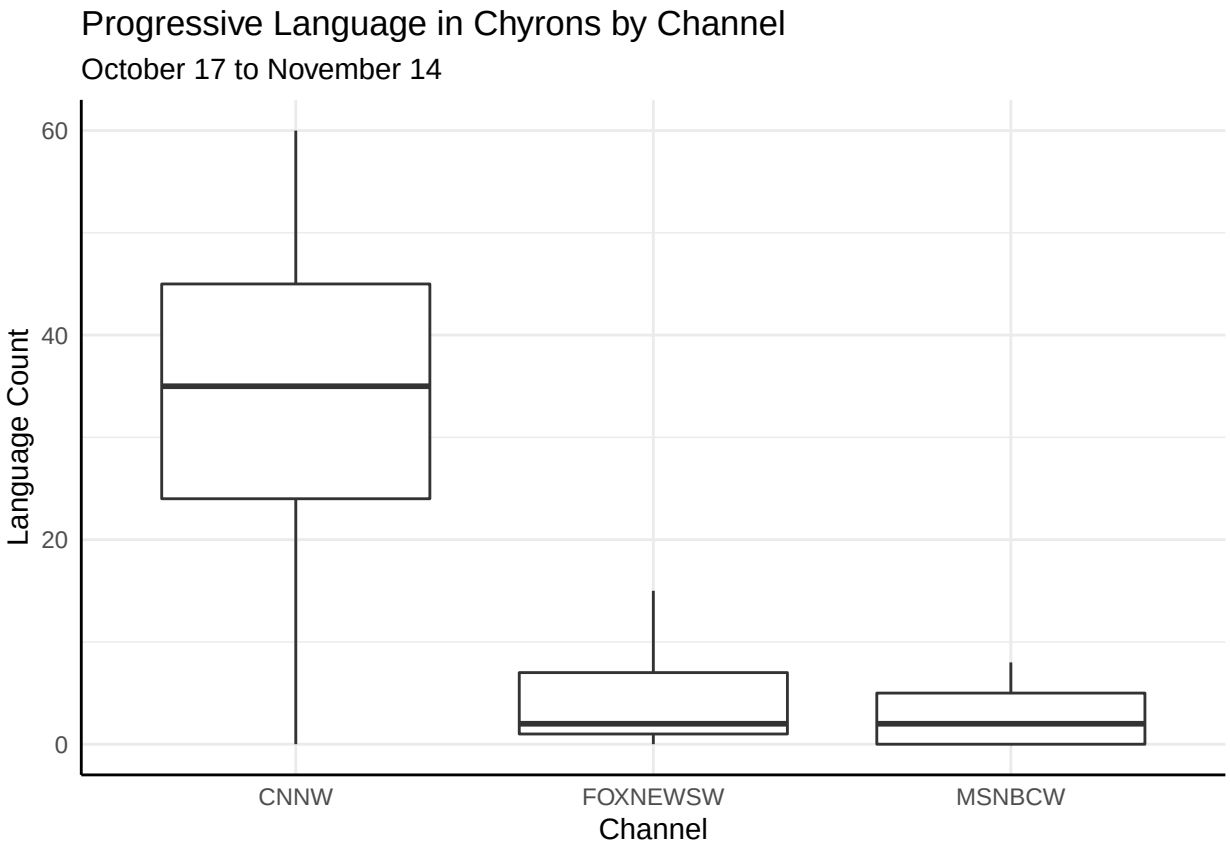
How much specific language is each channel using each day?

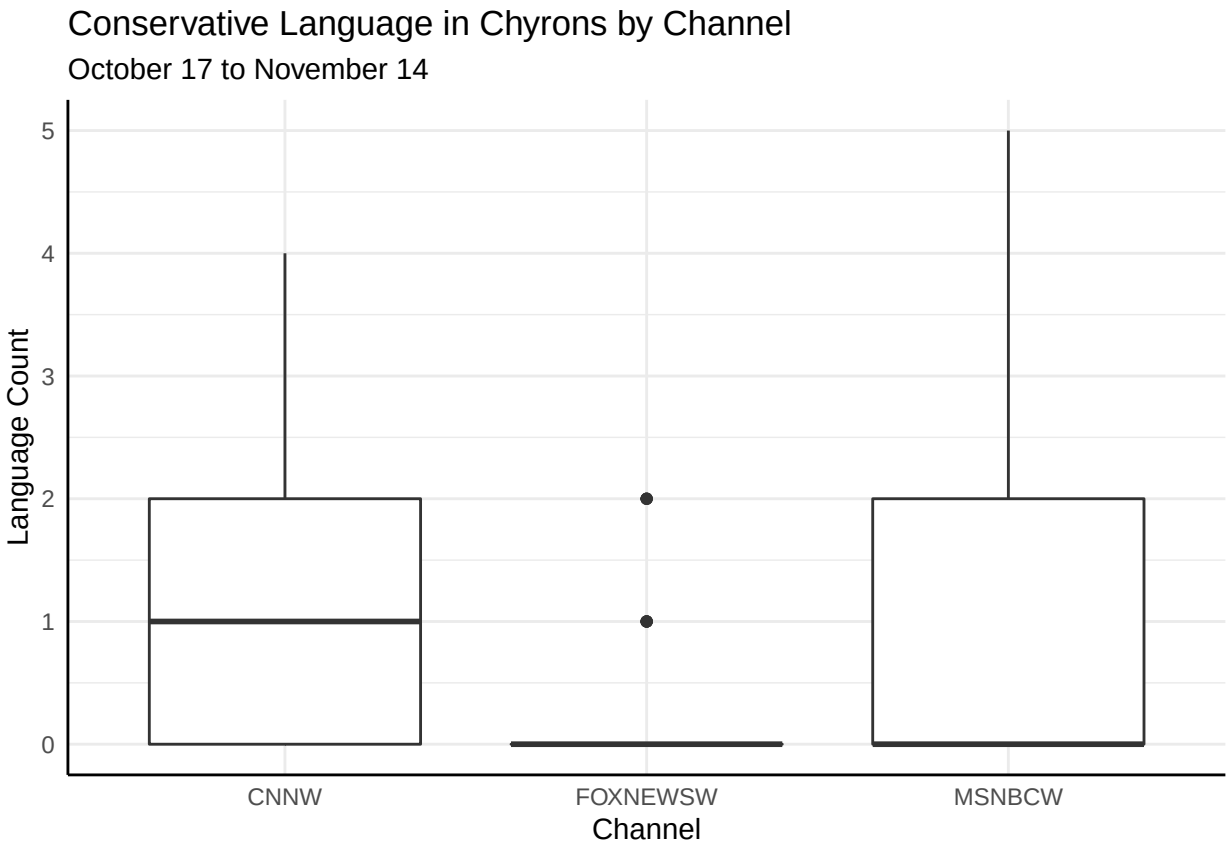
The histograms below visualize how much channels use the categorical language on a daily basis.


Populist Language in Chyrons by Channel
October 17 to November 14
Language Count
60
40
20
0
CNNW
FOXNEWSW
MSNBCW
Channel


Immigration Language in Chyrons by Channel
October 17 to November 14
Language Count
10
5
0
CNNW
FOXNEWSW
MSNBCW
Channel


Environmental Language in Chyrons by Channel
October 17 to November 14
Language Count
12.5
10.0
7.5
5.0
2.5
0.0
CNNW
FOXNEWSW
MSNBCW
Channel


Progressive Language in Chyrons by Channel
October 17 to November 14
Language Count
60
40
20
0
CNNW
FOXNEWSW
MSNBCW
Channel



How does usage vary across time?

The plots below visualize the change in daily language use for each channel.

Regression: Does Ideology have an affect on language usage?

The table below shows the results of five different regressions and the plot below shows the line of best fit for each regression.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sun, Dec 13, 2020 - 10:32:37

There does not appear to be a strong relationship between ideology and language usage.

	Dependent variable:				
	populism	immigration	environment	progressive	conservative
	(1)	(2)	(3)	(4)	(5)
as.factor(channel)FOXNEWSW	−14.810*** (1.300)	−0.414 (0.438)	−0.241 (0.242)	−14.362*** (1.779)	−0.431*** (0.155)
as.factor(channel)MSNBCW	−13.835*** (1.305)	−0.350 (0.440)	−0.084 (0.243)	−14.888*** (1.787)	0.006 (0.156)
as.factor(election)Pre-Election	3.998*** (1.096)	1.018*** (0.369)	0.070 (0.204)	2.469 (1.499)	0.085 (0.131)
as.factor(primetime)Primetime	−9.442*** (1.064)	−2.096*** (0.359)	−0.596*** (0.198)	−11.477*** (1.457)	−0.640*** (0.127)
Constant	19.860*** (1.262)	2.106*** (0.425)	0.755*** (0.235)	20.413*** (1.728)	0.871*** (0.151)
Observations	173	173	173	173	173
R^2	0.602	0.204	0.057	0.479	0.177
Adjusted R^2	0.592	0.185	0.035	0.466	0.158
Residual Std. Error (df = 168)	6.999	2.358	1.305	9.580	0.837
F Statistic (df = 4; 168)	63.461***	10.734***	2.546**	38.567***	9.057***

Note: *p<0.1; **p<0.05; ***p<0.01