Fox vs. Vox:

A linguistic assessment of subjectivity in U.S. news sources

Claire McLean
CAM685@pitt.edu
University of Pittsburgh

Table of CONTENTS

BACKGROUND

- Many of us have different perceptions of the level and direction of bias in different U.S. news sources
- Starting framework: **AllSides Media Bias Chart**
  - Assesses bias, not accuracy in reporting
  - Based on input from ‘ordinary’ Americans and experts combined
  - Not AI-based
- Selected one source from the ‘Left’ side and one from the ‘Right’ side
  - Vox (left)
  - Fox (right)

LITERATURE REVIEW

How can we operationalize bias in a linguistic way?

- **Linguistic Models for Analyzing and Detecting Biased Language (Recasens, et. al, 2013)**
  - Framing bias: subjective or one-sided terms reveal author's stance on a topic
    - Factive verbs
    - One-sided terms
    - Subjective intensifiers
      - a) Schnabel himself did the fantastic reproductions of Basquiat's work.
      - b) Schnabel himself did the accurate reproductions of Basquiat's work.

- **Intensification for discursive evaluation: a corpus-pragmatic view (Pan, 2021)**
  - Words like 'very', 'so', and 'really' add social and emotion-driven emphasis to utterances
  - Construction: Intensifier + adjective
  - Corpus: BNC Sampler
  - Tagset: CLAWS7 (degree adverbs / RG)

For the purposes of this project, I decided to analyze the top 4 degree adverbs (subjective intensifiers) from (Pan, 2021) within my own corpus. These were 'very', 'so', 'quite', and 'too'.

RESEARCH QUESTIONS

01 How do the frequencies of subjective intensifier constructions in the standard news articles compare to ‘non-subjective’ text?

02 Is there a significant difference in the prevalence of subjective intensifiers between the standard news articles and the ‘opinion’ articles from the same sources?

METHODS AND PROCEDURES

- Step 1: Identify key subjective intensifiers and select news sources for analysis
- Step 2: Create the following lists of data:
  - 25 Fox News URLs
  - 25 Fox News Opinion URLs
  - 25 Vox URLs
  - 25 Vox (The Big Idea) URLs
  - Brown Corpus (unbiased?)
- Step 3: Scrape websites
  - BeautifulSoup
- Step 4: Data cleaning
  - Built some cleaning into web scraper
  - Headline, (subheading), text, source
- Step 5: Linguistic analysis
- Step 6: Statistical analysis

THE DATA

25 of most recent articles were taken directly from the 'Politics' page of each news source.

(These were not chosen based on any criteria except for their ability to be scraped by the parser I created.)

25 'Opinion' articles were also taken from each of the same sources.

Fox News

Standard:

- 20,442 tokens
- 20,417 bigrams

Opinion:

- 25,385 tokens
- 25,360 bigrams

Vox

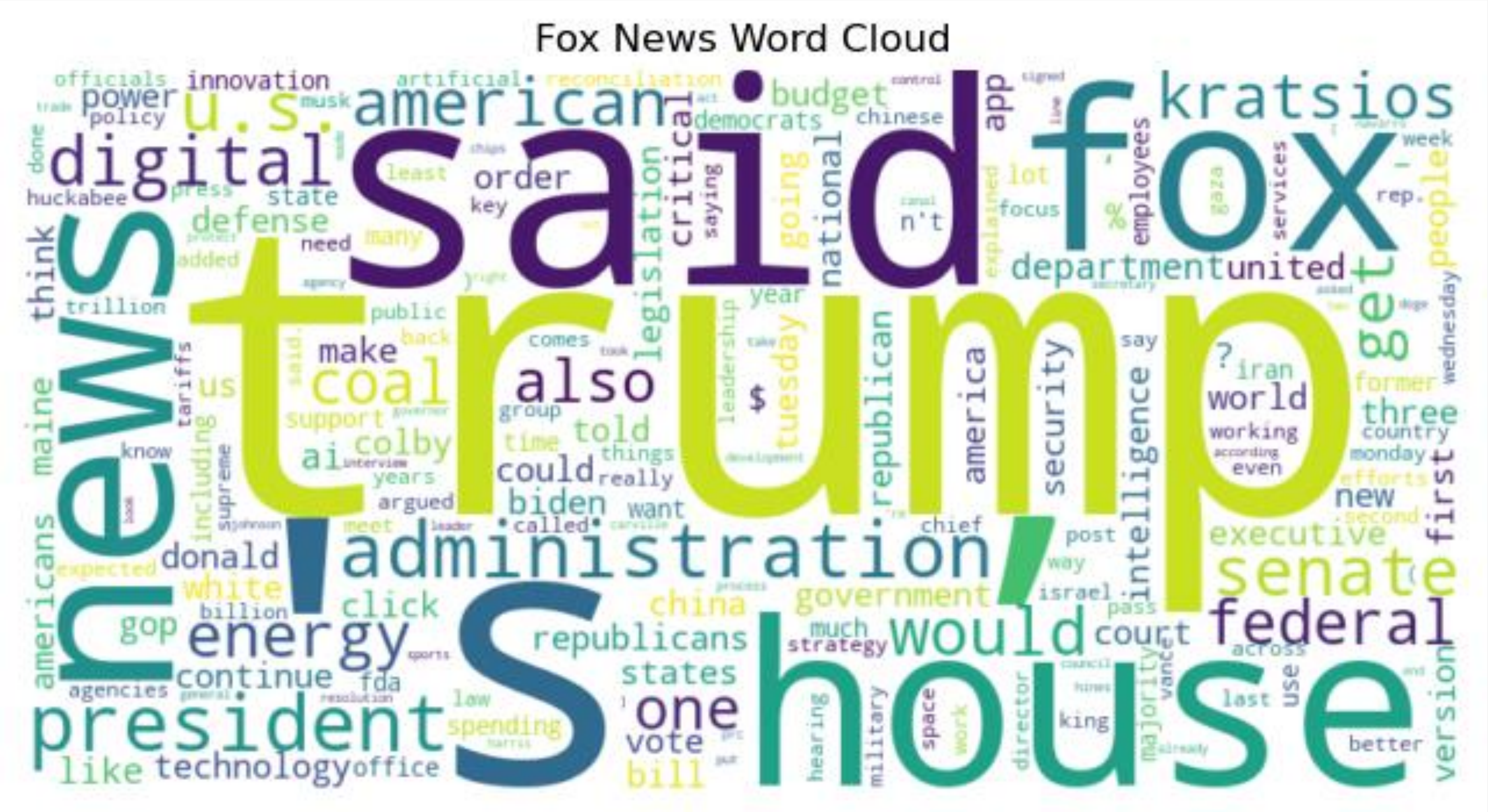
Standard:

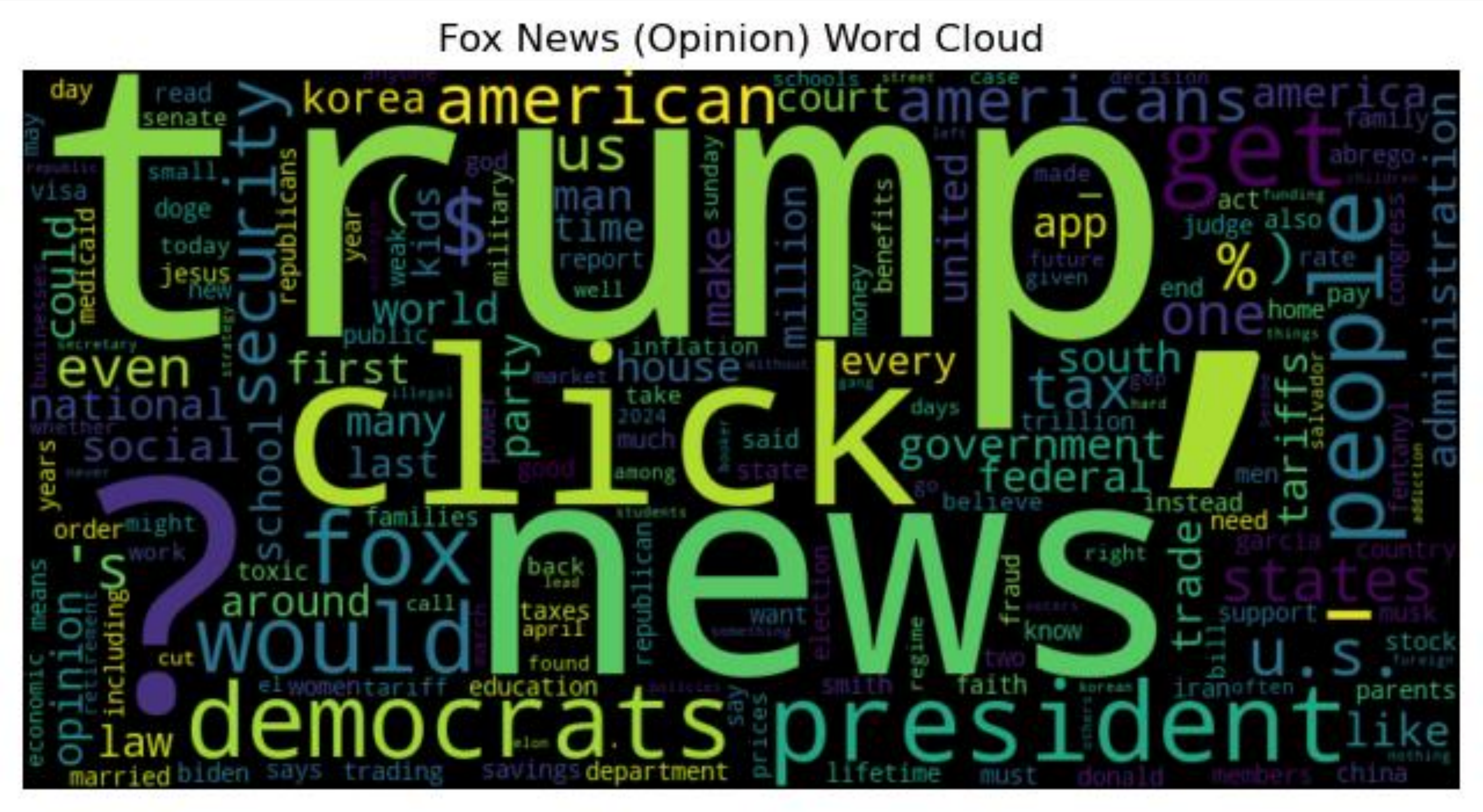
- 31,689 tokens
- 31,664 bigrams

Opinion:

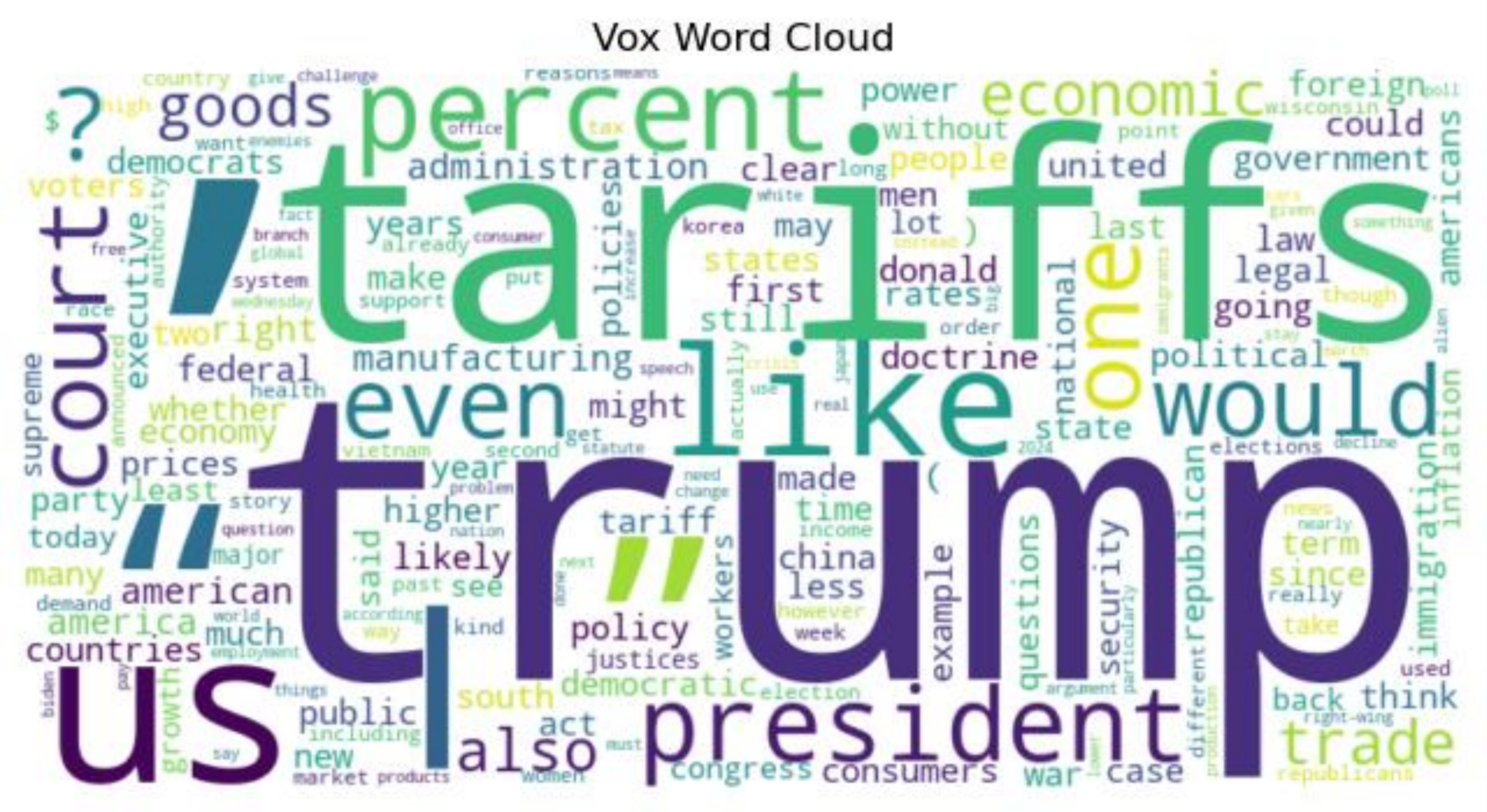
- 48,751 tokens
- 48,726 bigrams

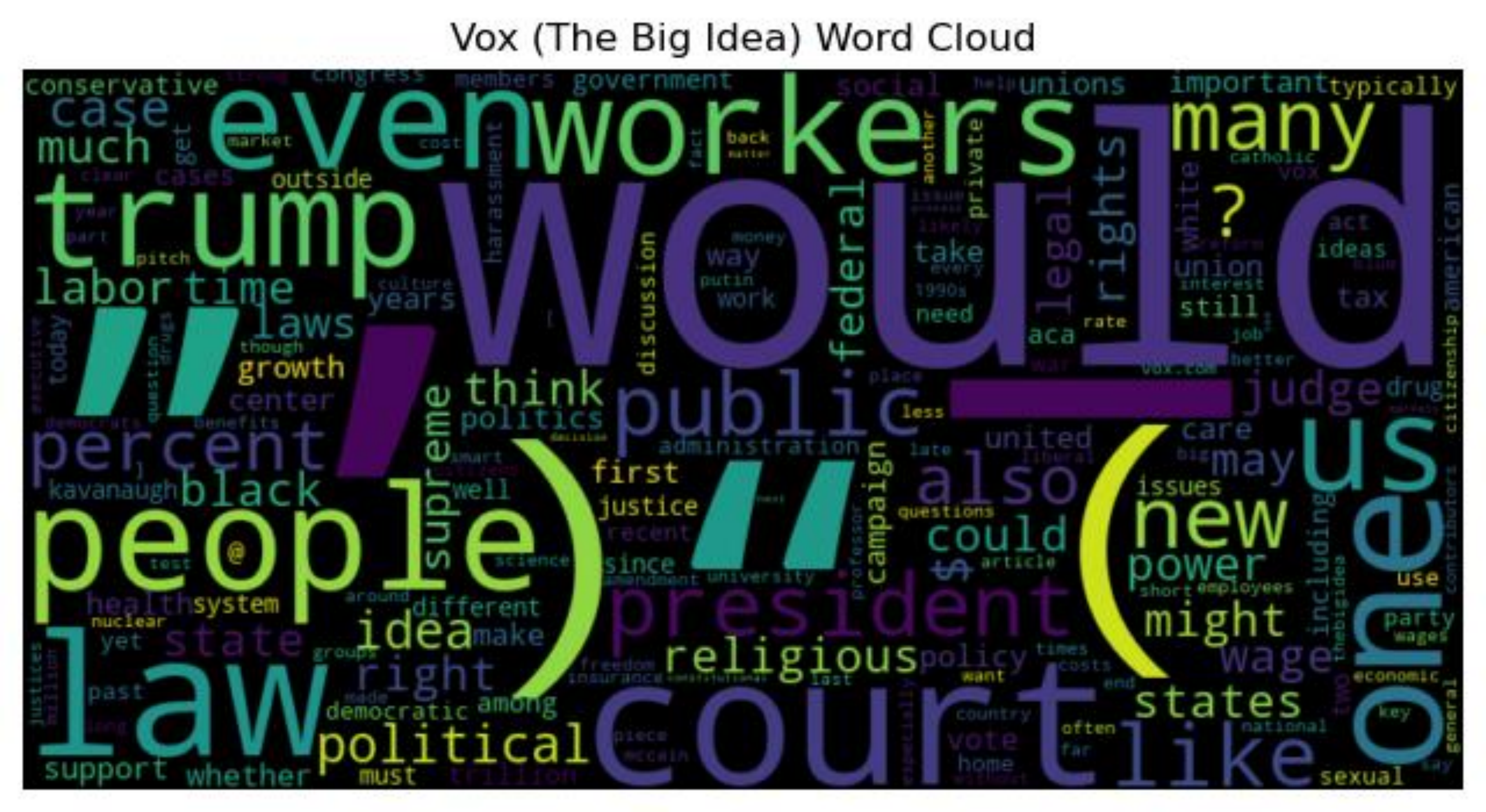
FOX NEWS


Fox News Word Cloud
trump
said
fox
news
house
administration
american
kratsios
digital
senate
federal
president
energy
would
coal
also
u.s.
intelligence
security
republican
department
united
government
republicans
executive
budget
democrats
legislation
critical
national
america
defense
order
people
get
technology
continue
donald
white
click
china
states
one
vote
bill
court
world
working
could
told
colby
biden
tuesday
like
innovation
officials
power
policy
huckabee
think
trillion
tariffs
maine
americans
gop
agencies
three
first
version
better
last
military
hearing
space
director
majority
law
spending
office


Fox News (Opinion) Word Cloud
trump
click
news
american
americans
democrats
president
would
fox
people
security
school
korea
court
us
app
get
america
administration
government
federal
tax
south
tariffs
trade
states
u.s.
even
national
social
first
many
last
world
million
united
every
house
party
kids
man
time
report
around
opinion
law
stock
support
like
china
department
education
families
inflation

VOX


Vox Word Cloud
tariffs
trump
percent
like
even
would
president
us
court
goods
economic
foreign
one
trade
also
policy
administration
government
democrats
manufacturing
political
legal
american
countries
federal
china
tariff
higher
likely
security
questions
republican
immigration
inflation
americans
consumers
congress
democratic
justices
workers
doctrine
donald
rates
first
states
economy
prices
party
today
many
public
new
act
war
case
think
back
time
less
made
year
might
state
national
going
law
last
lot
may
clear
power
without
people
united
could
voters
executive
supreme
policies
years
make
still
order
south
example
term
since
take
said
see
much
america
week
however


Vox (The Big Idea) Word Cloud
would
even
workers
many
trump
people
law
court
president
public
percent
one
us
like
new
also
could
power
might
states
political
religious
right
idea
black
think
politics
laws
time
labor
case
much
federal
legal
rights
supreme
judge
first
justice
growth
system
state
policy
wage
take
union
still
ideas
tax
care
may
issues
since
different
make
center
support
whether
years
administration
conservative
government
unions
important
typically
social
members
congress
outside
cases
campaign
health
kavanaugh
vote
home
often
sexual
must
way
work
need
party
use
democratic
among
act

Bigram Frequency Distributions

Fox News bigram frequency distribution (top 10)

{('&', 'nbsp'): 207, ('nbsp', ';'): 207, ('.', '‘'): 145, (',', '"'): 137, ('’', 's'): 91, ('of', 'the'): 88, (',', 'and'): 80, ('fox', 'news'): 79, ('in', 'the'): 72, ('to', 'the'): 71}

Fox News Opinion bigram frequency distribution (top 10)

{('&', 'nbsp'): 675, ('nbsp', ';'): 675, ('’', 's'): 179, ('in', 'the'): 84, ('’', 't'): 76, (',', 'and'): 76, ('of', 'the'): 68, (',', 'the'): 63, ('click', 'here'): 61, ('.', 'the'): 58}

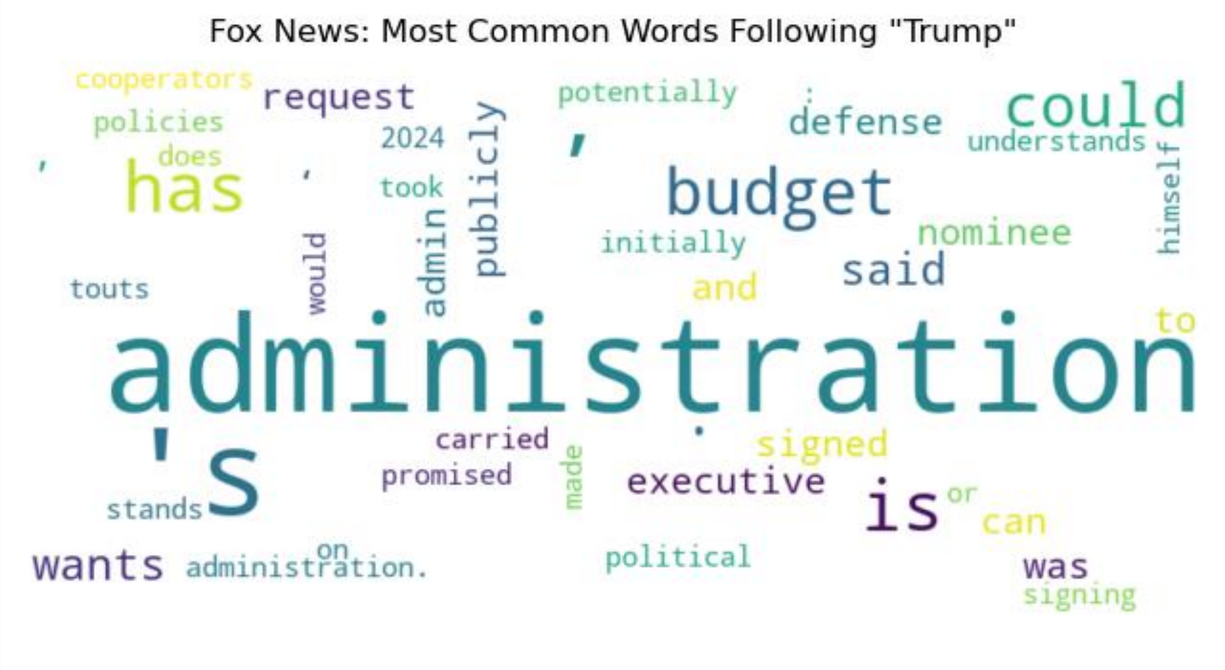
Vox bigram frequency distribution (top 10)

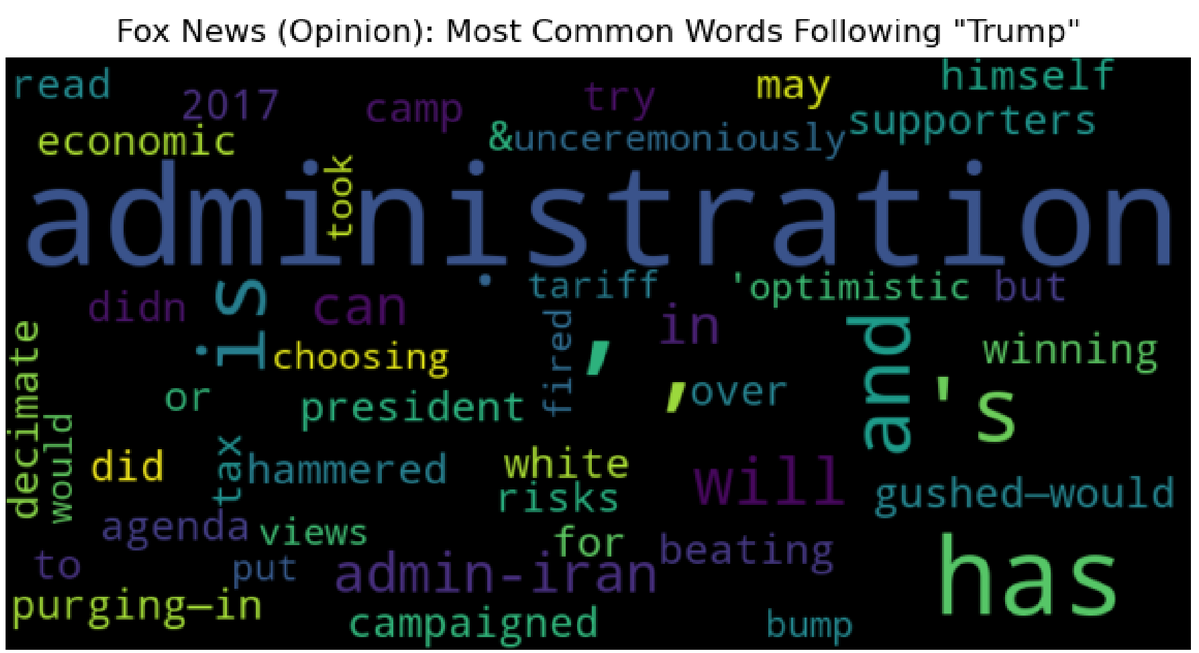
{('’', 's'): 492, (',', 'and'): 159, ('in', 'the'): 130, ('trump', '’'): 127, ('of', 'the'): 121, (',', 'the'): 112, ('.', 'the'): 98, ('’', 't'): 82, ('.', 'and'): 74, ('.', 'but'): 72}

Vox (The Big Idea) bigram frequency distribution (top 10)

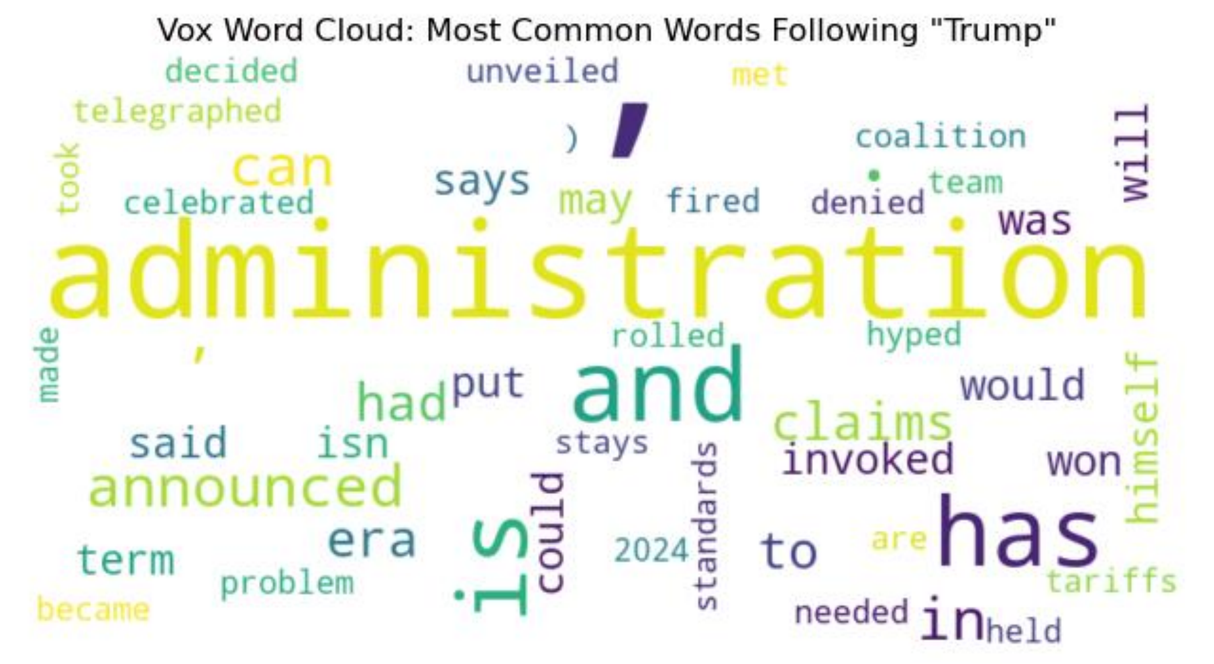
{('’', 's'): 425, ('of', 'the'): 318, (',', 'and'): 245, ('in', 'the'): 194, ('.', 'the'): 189, (',', 'the'): 183, ('to', 'the'): 120, ('.', 'but'): 107, ('’', 't'): 103, ('on', 'the'): 100}

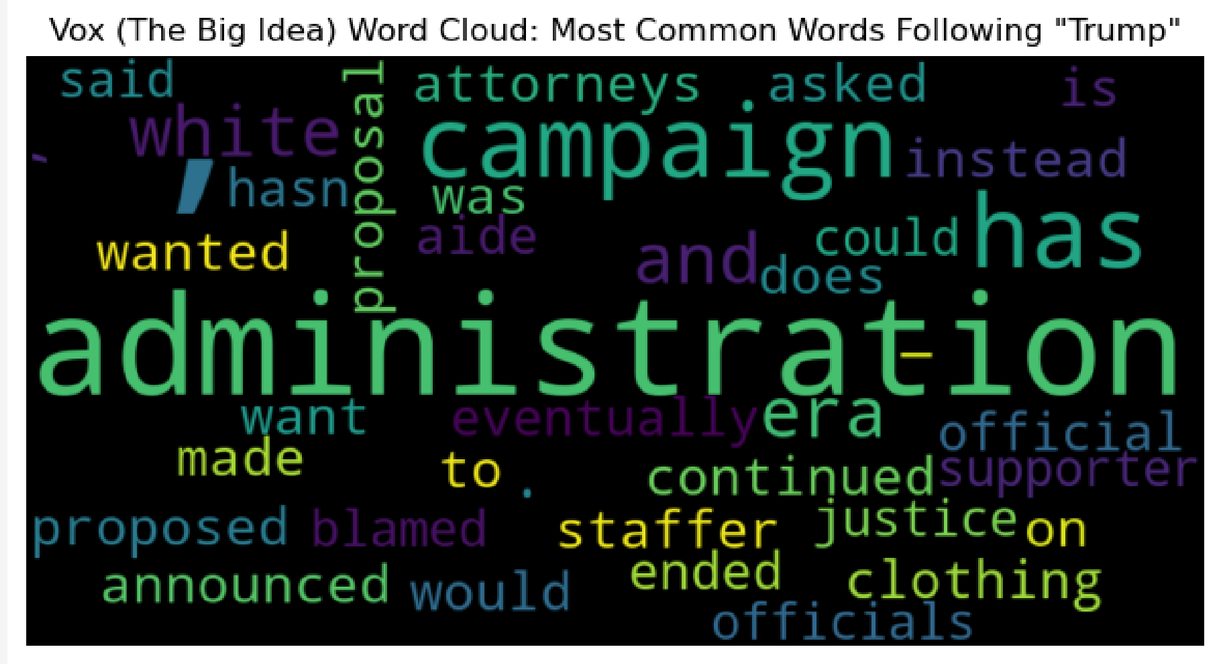
FOX NEWS CFD: WORDS FOLLOWING ‘TRUMP’





VOX CFD: WORDS FOLLOWING ‘TRUMP’





NEXT STEPS

TAG BIGRAMS

Necessary to identify subjective intensifiers that are followed by **adjectives**

IDENTIFY SUBJECTIVE CONSTRUCTIONS IN ALL CORPORA

This will be accomplished through conditional frequency distributions. Finding ADJ words that follow each of the 4 intensifiers

COMPARE SUBJECTIVE CONSTRUCTION FREQUENCIES BETWEEN TEXTS

REFERENCES

Recasens, M., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2013). Linguistic Models for Analyzing and Detecting Biased Language. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 1650–1659

Pan, Y. (2022). Intensification for discursive evaluation: a corpus-pragmatic view. Text & Talk, 42(3), 391-417.