

[코퍼스언어학]기말과제

트럼프와 바이든 취임연설문 비교



| | |
|-------|------------|
| ▶ 학과 | 경제학과 |
| ▶ 학번 | 20190543 |
| ▶ 성명 | 박희진 |
| ▶ 과목 | 코퍼스언어학 |
| ▶ 제출일 | 2021.12.24 |

<목차>

| | |
|----------------|----|
| I. 서론..... | 3 |
| II. 본론..... | 3 |
| 1. 빈도분석..... | 3 |
| 2. n-gram..... | 6 |
| 3. 연어 | 8 |
| 4. 키워드 분석..... | 10 |
| III. 결론..... | 12 |

I. 서론

수업시간에 Washington(1789)부터 시작해서 Obama(2009)까지 미국대통령 취임연설문 텍스트에 대해 다루는 내용을 흥미롭게 들었지만 Obama 이후에 취임된 Trump(2017)와 Biden(2021) 관해서는 수업시간에서 아예 다루지 않았던 점이 아쉬웠다. 그러던 와중 두 대통령 연설문 사이에는 어떠한 차이가 있을지, 각자 추구하고자 하는 정책의 내용이 텍스트 분석 결과에 잘 나타나 있는지, 두 대통령이 공통적으로 쓴 어휘는 무엇인지, 그 어휘를 왜 많이 사용했는지, 그 어휘랑 같이 사용한 어휘쌍은 무엇이 있는지 등 궁금증이 끊임없이 생겨 트럼프와 바이든의 미국대통령 취임연설문을 관찰 데이터로 선정하게 되었다.

Trump 와 Biden 의 취임연설문 텍스트 가공 및 분석에 있어서 다음과 같은 순서에 의해 작성할 것이다. 첫 번째로는 빈도분석을 통해 어떤 단어가 고빈도로 나타나는지 살펴볼 것이며 두 번째로는 bi-gram 을 통해 두 단어의 연쇄로 어떤 단어쌍들이 많이 나타나는지 확인할 것이다. 세 번째로는 연어를 통해 중심어 주변에 같이 사용되는 언어쌍을 구할 것이고 네 번째로는 키워드 분석을 통해 positive keyword 가 무엇이 나오는지 확인해 봄으로써 트럼프와 바이든의 연설문에는 어떠한 차이가 있는지 알아낼 것이다.

II. 본론

1. 빈도분석

□ 이름

2017_Trump

2021_Biden

stop

2017_Trump - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

Chief Justice Roberts, President Carter, President Clinton. Every four years, we gather on these steps to carry out our duty. Today's ceremony, however, has very special meaning. For too long, a small group in our nation's capital has stood in the way of progress. What truly matters is not which party controls our government, but whether we are young and beautiful students deprived of all knowledge. We are one nation and their pain is our pain. Their suffering is our suffering. But that is the past, and now we are looking only to the future. Every decision on trade, on taxes, on immigration, on health care will be made with the best interests of the American people in mind. We will follow two simple rules: buy American, and invest in American jobs. At the bedrock of our politics will be a total allegiance to the Constitution. Finally, we must think big and dream even bigger. In the words of our founding fathers, We stand at the birth of a new millennium, ready to embrace the challenges and opportunities of the 21st century. So to all Americans, in every city near and far, small

2021_Biden - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

Chief Justice Roberts, Vice-President Harris, Speaker of the House Nancy Pelosi, and I. This is America's day. This is democracy's day. America has learned again that democracy is precious, and we must protect it. As we look ahead in our uniquely American way, I've just taken a sacred oath each of those patriots who have served this country. We'll press forward with speed and urgency for victory. Millions of jobs have been lost. Hundreds of thousands of Americans have been left behind. To overcome these challenges, to restore the soul of this country, my whole soul is in it today, on this January day. With unity we can do great things, important things that will last for generations. I know speaking of unity can sound to some like a cliché. Through civil war, the Great Depression, World War

stop - Windows 메모장

파일(F) 편집(E) 서식(O)

a's
able
about
above
according
accordingly
across
actually
after
afterwards
again
against

동일한 폴더 안에 2017_Trump.txt(트럼프 취임연설문 텍스트), 2021_Biden(바이든 취임연설문 텍스트), stop.txt(english stop words 불용어가 들어간 텍스트)가 있다..

트럼프와 바이든의 취임연설문 텍스트는 오직 대사로만 이뤄져 있다(괄호 안에 지문이 들어간 경우가 없다). 연설문 텍스트를 잘 분석하기 위해 소문자로 변환하고 gsub 함수에 정규표현을 적용하여 양 사이드에 있는 문장부호를 제거했으며 빈문자열을 제거했다. Trump 와 Biden 이 공통적으로 많이 사용한 어휘가 무엇인지 알고 싶어 빈도내림차순의 dataframe 을 만들었다.

```
> #트럼프 취임연설문 텍스트: 2017_Trump.txt
> Trump<-scan(file='2017_Trump.txt',what='',quote=NULL)
Read 1464 items
> Trump<-gsub('^[:punct:]]+|[:punct:]]+$','',tolower(Trump))
> Trump<-Trump[nchar(Trump)>0]
> head(Trump)
[1] "chief"      "justice"    "roberts"    "president"  "carter"     "president"
> #바이든 취임연설문 텍스트: 2021_Biden.txt
> Biden<-scan(file='2021_Biden.txt',what='',quote=NULL)
Read 2427 items
> Biden<-gsub('^[:punct:]]+|[:punct:]]+$','',tolower(Biden))
> Biden<-Biden[nchar(Biden)>0]
> head(Biden)
[1] "chief"      "justice"    "roberts"    "vice-president"
[5] "harris"     "speaker"

> Trump.freq<-data.frame(sort(table(Trump),decreasing=T))
> Biden.freq<-data.frame(sort(table(Biden),decreasing=T))

> head(Trump.freq,15)    > head(Biden.freq,15)
      Trump Freq      Biden Freq
1      and    76    1      the   110
2      the    73    2      and   100
3      our    49    3      we    78
4      of     48    4      of    77
5      we     48    5      to    63
6      will   43    6      a     48
7      to     37    7      in    44
8      is     20    8      our   44
9  america   19    9      i     38
10     a      15   10     this   33
11    for     15   11    for    29
12   you      15   12     us    25
13   all      14   13   will    25
14    in      14   14     my    24
15    be      13   15     as    23
```

왼쪽은 Trump 연설문에 나타나는 어휘의 빈도내림차순으로 만들어진 dataframe, 오른쪽은 Biden 연설문에 나타나는 어휘의 빈도내림차순으로 만들어진 dataframe 이다. Trump 와 Biden 모두 고빈도 어휘에 비슷한 function 어들이 많이 포함되어 있음을 알 수 있다. function 어들이 많으면 해석하기 어렵기 때문에 불용어를 제거하여 고빈도 어휘에 나타나는 contents word 에 대해 더 자세히 비교해보았다.

```
> #stop words list에 포함된 어휘 제거해주기
> #stop words list 텍스트: stop.txt
> stop<-scan(file='stop.txt',what='',quote=NULL)
Read 571 items
> filter.Trump<-Trump[!(Trump%in%stop)]
> filter.Trump.freq<-data.frame(sort(table(filter.Trump),decreasing=T))
> filter.Biden<-Biden[!(Biden%in%stop)]
> filter.Biden.freq<-data.frame(sort(table(filter.Biden),decreasing=T))
```

%in%와 !를 통해 불용어를 제거했으며 빈도내림차순으로 구성된 dataframe 을 만들었다.

| > head(filter.Trump.freq,15) | | | > head(filter.Biden.freq,15) | | |
|------------------------------|--------------|------|------------------------------|--------------|------|
| | filter.Trump | Freq | | filter.Biden | Freq |
| 1 | america | 19 | 1 | america | 17 |
| 2 | american | 11 | 2 | nation | 12 |
| 3 | people | 10 | 3 | americans | 11 |
| 4 | country | 9 | 4 | democracy | 11 |
| 5 | back | 6 | 5 | people | 11 |
| 6 | great | 6 | 6 | story | 9 |
| 7 | make | 6 | 7 | unity | 8 |
| 8 | nation | 6 | 8 | world | 8 |
| 9 | world | 6 | 9 | american | 7 |
| 10 | dreams | 5 | 10 | day | 7 |
| 11 | president | 5 | 11 | days | 7 |
| 12 | protected | 5 | 12 | history | 7 |
| 13 | americans | 4 | 13 | stand | 7 |
| 14 | bring | 4 | 14 | today | 7 |
| 15 | citizens | 4 | 15 | war | 7 |

왼쪽은 불용어를 제거한 Trump 의 빈도내림차순 dataframe, 오른쪽은 불용어를 제거한 Biden 의 빈도내림차순 dataframe 이다. 불용어를 제거함으로써 어떤 contents word 가 많이 사용되었는지 확인이 가능해졌다. 불용어를 제거한 Trump 와 Biden 에서 공통적으로 고빈도에 나타나는 어휘가 무엇인지 알기 위해 intersect 를 이용해 고빈도어휘 15 개를 비교해보았다.

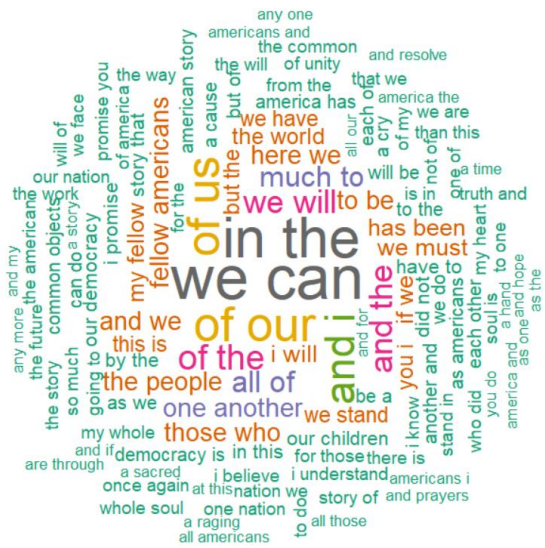
```
> head_t<-head(filter.Trump.freq$filter.Trump,15)
> head_b<-head(filter.Biden.freq$filter.Biden,15)
> intersect(head_t,head_b)
[1] "america" "american" "people" "nation" "world" "americans"
```

공통적으로 고빈도에 나타나는 어휘에는 america, american, people, nation, world, americans 가 있음을 알 수 있다. 여기서 특히 america 와 관련된 어휘들이 많이 나타나는 것을 찾아볼 수 있다. 현재 분석하고자 하는 텍스트는 미국 대통령 취임연설문이기 때문에 미국에 대한 정책이나 미국인에게 하고자 하는 말이 담겨져 있을 것이다. america 와 관련된 어휘에 대해서는 연어분석할 때 더 자세히 다룰 예정이다.

2. n-gram

Trump 와 Biden 은 어떤 bi-gram 을 많이 사용했을까? paste 함수를 사용하여 앞뒤에 연달아 있는 두 어휘들을 묶어 bi-gram 으로 만들었으며 빈도의 차이를 시각화 정보로 나타내주는 wordcloud 를 이용하여 확인해보았다. wordcloud 는 많은 어휘들을 한 번에 볼 수 있어서 관찰하는 어휘빈도를 분석하기 용이하게 해준다. 이제부터는 불용어가 제거되지 않은 텍스트로 다룰 것이다. 그 이유는 불용어를 제거한 채 n-gram 이나 연어분석을 하면 결과가 왜곡될 가능성이 매우 크기 때문이다.

```
> Trump_bi.grams<-paste(Trump[1:length(Trump)-1],Trump[2:length(Trump)])
> Trump_bi.Freq<-data.frame(sort(table(Trump_bi.grams),decreasing=T))
> Biden_bi.grams<-paste(Biden[1:length(Biden)-1],Biden[2:length(Biden)])
> Biden_bi.Freq<-data.frame(sort(table(Biden_bi.grams),decreasing=T))
> library(wordcloud)
Loading required package: RColorBrewer
> wordcloud(Trump_bi.Freq$Trump_bi.grams,Trump.freq$Freq,
+ scale=c(4,1),min.freq=4,max.word=120,random.order=F,rot.per=0.4,
+ colors=brewer.pal(8,"Dark2"))
> wordcloud(Biden_bi.Freq$Biden_bi.grams,Biden.freq$Freq,
+ scale=c(4,1),min.freq=4,max.word=120,random.order=F,rot.per=0.4,
+ colors=brewer.pal(8,"Dark2"))
```

Trump에서 살펴볼 만한 고빈도 bi-gram으로 we will이 있다. we will에서 will은 ‘할 것이다’라는 뜻으로 아직 사실이 확인되지 않았지만 미래에 확실한 사실로 나타날 것이라는 화자의 확신과 강력한 의지를 담고 있는 표현이다. we와 will을 함께 사용함으로써 Trump는 국민들과 함께 해나갈 것이라는 확고한 의지를 표현했다고 유추할 수 있다. 또한 of america, make america, american workers, bless america, america first 등 america와 관련된 어휘가 포함되는 bi-gram이 많이 나타나는 것을 확인할 수 있다.

Biden은 we will도 사용했지만 이보다 we can을 더 많이 사용했다. can은 ‘할 수 있다’라는 의미로 기본적으로 능력을 나타내며 용기를 심어주고 복돋아주는 표현이다. we와 can을 함께 사용함으로써 Biden은 국민들에게 할 수 있다는 용기를 복돋아주었다고 유추할 수 있다. 또한 고빈도 bi-gram인 fellow americans, my fellow에서 볼 수 있듯이 국민을 가리킬 때 fellow를 많이 사용했음을 알 수 있다. Biden에서도 마찬가지로 american story, fellow american, america has 등 america와 관련된 어휘를 포함되는 bi-gram이 많이 나타나는 것을 확인할 수 있다.

america관련 어휘는 Trump와 Biden에서 모두 빈도가 높게 나타나고 bi-gram에서도 고빈도로 나타난다. 이런 결과를 보고 america관련 어휘는 어떤 어휘들과 함께 빈번하게 사용되는지에 대한 궁금증이 생겼다. 또한 Trump와 Biden이 각자 추진하고자 하는 미국정책이나 미국 국민들에게 말하고자 하는 이야기가 america관련 어휘 근처에서 연어분석을 통해 더 명확히 관찰할 수 있다고 생각하여 america관련 어휘를 중심으로 설정하여 연어분석을 진행했다.

3. 연어

America와 관련된 어휘들을 앞으로 중심어라고 지칭하겠다. 앞에서 볼 수 있듯이 중심어는 Trump, Biden 모두에게 많이 나타나는 어휘이다. 공기빈도만 보는 bi-gram보다 span구간을 설정하여 관찰빈도와 기대빈도를 통해 계산식이 나온 관점에서 연어분석을 하면 더 효과적으로 파악할 수 있다. 따라서 중심어의 좌우 각각 두 구간씩, 총 4구간에 대해 살펴보았다. MI-score가 높을 때 텍스트 데이터를 해석하기 용이한 content word가 많이 출력되기 때문에 t-score가 아닌 MI-score을 이용해 살펴보았으며 공기빈도가 낮은 MI-score의 단점을 보완하기 위해 최소공기빈도를 2보다 크게 제한했다.

```
> #Trump에서 연어추출
> #1단계:공기어 추출
> node<-'america.*'
> index<-grep(node,Trump)
> span<-vector()
> for(i in index)
+ {
+ span<-c(span,c((i-2):(i-1),(i+1):(i+2)))
+ }
> span<-span[span>0&span<length(Trump)]
> crc<-Trump[span]
> #2단계:공기어 데이터프레임
> Freq.span<-sort(table(crc),decreasing=T)
> Freq.all<-table(Trump)
> Freq.co<-data.frame(W1=vector(),W2=vector(),W1W2=vector(),N=vector())
> Freq.co<-data.frame(t(apply(names(Freq.span),
+ function(x){c(length(index),Freq.all[x],Freq.span[x],length(Trump))})))
> colnames(Freq.co)<-c('W1','W2','W1W2','N')
> #3단계:연어계산
> collocates<-data.frame(Freq.co,
+ MI=log2((Freq.co$W1W2*Freq.co$N)/(Freq.co$W1*Freq.co$W2)))
> MI.sort_t<-collocates[order(collocates$MI,decreasing=T),]
```

위의 코드는 중심어를 america관련 어휘로 설정하여 Trump에서 공기어를 추출하고 데이터 프레임 만들어 MI-score 값을 추가한 코드이다.


```

> #Biden에서 단어추출
> #1단계:공기어 추출
> node<-'america.*'
> index<-grep (node,Biden)
> span<-vector ()
> for (i in index)
+ {
+ span<-c (span,c ((i-2):(i-1),(i+1):(i+2)))
+ }
> span<-span[span>0&span<length(Biden)]
> crc<-Biden[span]
> #2단계:공기어 데이터프레임
> Freq.span<-sort (table (crc),decreasing=T)
> Freq.all<-table (Biden)
> Freq.co<-data.frame (W1=vector (),W2=vector (),W1W2=vector (),N=vector ())
> Freq.co<-data.frame (t (sapply (names (Freq.span),
+ function (x) {c (length (index),Freq.all [x],Freq.span [x],length (Biden))})))
> colnames (Freq.co)<-c ('W1','W2','W1W2','N')
> #3단계:연어계산
> collocates<-data.frame (Freq.co,
+ MI=log2 ((Freq.co$W1W2*Freq.co$N)/(Freq.co$W1*Freq.co$W2)))
> MI.sort_b<-collocates[order (collocates$MI,decreasing=T),]

```

위의 코드는 중심어를 america관련 어휘로 설정하여 Biden에서 공기어를 추출하고 데이터 프레임 만들어 MI-score 값을 추가한 코드이다.

최소 공기빈도를 2 보다 크게 제한하고 MI-score 를 기준으로 내림차순 정렬하면 다음과 같은 결과가 나온다. 왼쪽은 Trump, 오른쪽은 Biden 에 대한 내용이다.

```

> head(MI.sort_t[MI.sort_t$W1W2>2,],10) > head(MI.sort_b[MI.sort_b$W1W2>2,],10)

```

| | W1 | W2 | W1W2 | N | MI | | W1 | W2 | W1W2 | N | MI |
|---------|----|----|------|------|----------|---------|----|----|------|------|----------|
| workers | 35 | 2 | 3 | 1463 | 5.970394 | fellow | 37 | 6 | 6 | 2417 | 6.029548 |
| make | 35 | 6 | 5 | 1463 | 5.122397 | story | 37 | 9 | 3 | 2417 | 4.444586 |
| first | 35 | 4 | 3 | 1463 | 4.970394 | has | 37 | 10 | 3 | 2417 | 4.292583 |
| again | 35 | 9 | 5 | 1463 | 4.537434 | my | 37 | 24 | 5 | 2417 | 3.766514 |
| great | 35 | 6 | 3 | 1463 | 4.385431 | i | 37 | 38 | 7 | 2417 | 3.588976 |
| is | 35 | 20 | 5 | 1463 | 3.385431 | america | 37 | 17 | 3 | 2417 | 3.527048 |
| all | 35 | 14 | 3 | 1463 | 3.163039 | all | 37 | 20 | 3 | 2417 | 3.292583 |
| will | 35 | 43 | 8 | 1463 | 2.959166 | is | 37 | 21 | 3 | 2417 | 3.222193 |
| of | 35 | 48 | 7 | 1463 | 2.607823 | as | 37 | 23 | 3 | 2417 | 3.090949 |
| and | 35 | 76 | 11 | 1463 | 2.596935 | that | 37 | 23 | 3 | 2417 | 3.090949 |

Trump는 중심어 근처에서 workers, make, first 등이 MI-score가 높게 나타나 있다. workers 가 높게 나온 결과를 통해 트럼프가 미국의 노동자나 실업자를 주요대상으로 여기고 있다는 사실을 알 수 있다. make의 값이 높게 나온 것은 미국을 어떤 방향으로 만들지에 대해 이야기를 하는 것으로 유추할 수 있고, first를 통해 미국 우선주의(American First)를 강조한 트럼프의 정책을 엿볼 수 있다.

Biden은 중심어 근처에서 fellow, story 등이 MI-score이 높게 나왔다. fellow의 값이 높은 것을 통해 Biden이 미국 자국민을 지칭할 때 fellow라는 어휘를 많이 사용되었음을 알 수 있다. 또한 story를 통해 미국의 역사와 관련된 이야기를 했음을 유추할 수 있다.

4.키워드 분석

□ 이름

2017_Trump
2021_Biden
stop

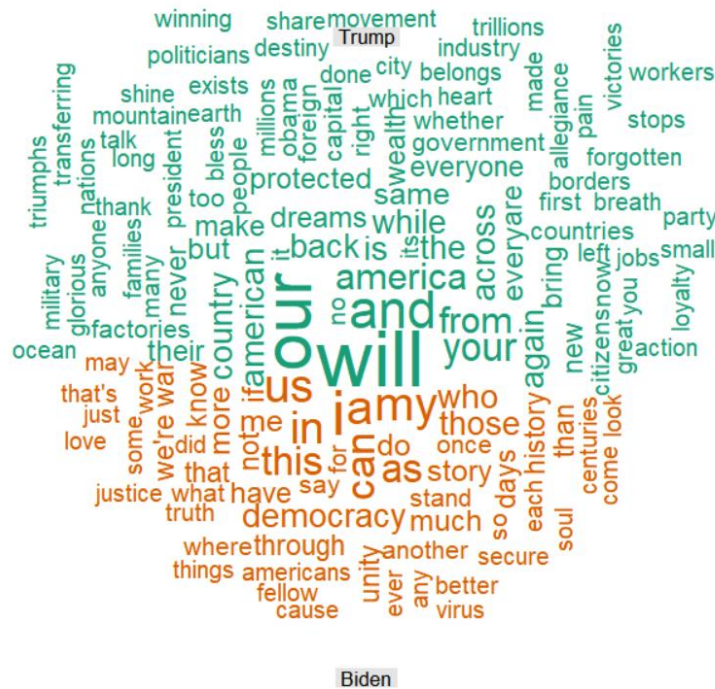
이번에는 파일이름을 직접 넣지 않고 for 문과 if 문을 이용해 원하는 파일만 읽어보려고 한다. 현재 Trump 와 Biden 의 연설문이 담긴 파일 이름은 취임연도_대통령이름.txt 이며 파일명에 숫자가 포함되어 있고 stop.txt 파일은 파일명에 숫자가 포함되어 있지 않은 것을 확인할 수 있다. 취임연설문의 텍스트만 읽어오려고 할 경우 for 문을 통해 현재 폴더 안에 있는 파일을 하나씩 반복시키고 if 문을 이용해 파일이름에 숫자가 포함되어 있는지 여부를 확인함으로써 숫자가 포함되어 있으면 파일을 읽어오도록 한다. 앞에서 언급했듯이 연설문 텍스트를 잘 분석하기 위해 소문자로 변환하고 문장부호와 빈문자열을 제거했다. column 명은 숫자 없이 대통령 이름만 나오게 substring 함수를 이용해서 정해주었다.

```
> TDM<-data.frame(words=vector())
> for (i in list.files())
+ {
+   if (length(grep('[:digit:]',i))>0)
+   {
+     file<-scan(file=paste('./',i,sep='/'),what='',quote=NULL)
+     file<-gsub('^[:punct:]]+|[:punct:]]+$','',tolower(file))
+     file<-file[nchar(file)>0]
+     TDM<-merge(TDM,data.frame(table(file)),by.x='words',by.y='file',all=T)
+     colnames(TDM)[length(TDM)]<-substring(gsub('[.].+?$', '',i),6)
+   }
+ }
Read 1464 items
Read 2427 items
```

기본적인 특성이 배타적인 comparison cloud 는 두 개의 text 만 비교하는 경우에는 큰 문제가 없기 때문에 comparison cloud 를 이용해 키워드 분석을 진행했다. comparison cloud를 이용해 keyword분석을 할 때 반드시 dataframe의 행명은 어휘가 되어야 한다. 또한

현재 읽어온 파일 이름은 숫자로 시작하기 때문에 dataframe 으로 만들면 column 명에 자동으로 X 가 붙어서 이를 gsub 함수를 이용해 없앴다.

```
> TDM<-data.frame(row.names=TDM$words,TDM[2:length(TDM)] )
> TDM[is.na(TDM)]<-0
> colnames(TDM)<-gsub('X','',colnames(TDM))
> comparison.cloud(TDM,random.order=FALSE,
+ scale=c(4,1),rot.per=.4,max.words=150,
+ colors=brewer.pal(8,'Dark2'),title.size=1.1)
```



Trump의 positive keyword는 무엇이 있을까? comparison cloud에서 살펴보았을 때 will, protected, jobs, workers를 찾아볼 수 있다. Trump는 will을 통해 확고한 의지를 나타내는 조동사를 많이 사용했고 America First를 정책핵심으로 강조한 만큼 트럼프는 미국의 이익을 우선시한다는 background를 가지고 있다면 protected라는 어휘는 미국을 보호하겠다는 의미로 사용되었음을 짐작할 수 있다. 또한 jobs와 workers를 통해 근로자와 실업자, 이에 대한 일자리 관련해서 이야기를 많이 한 것으로 보인다.

Biden의 연설문에서는 can, unity, uniting, democracy, virus 가 positive keyword로 나온 것을 확인할 수 있다. Biden은 can을 통해 할 수 있다는 능력을 나타내며 용기를 북돋우는 조동사를 많이 사용했다. 통합과 관련된 어휘들인 unity, uniting과 민주주의를 뜻하는 democracy가 positive keyword라는 사실을 통해 Biden은 민주주의와 결합, 통합, 화합을 연설문에서 매우 강조했음을 알 수 있다. 또한 2019년에 시작된 COVID-19은 Trump가 대통령 취임 이후에 발

생했기 때문에 Trump의 연설문에는 바이러스 관련 내용이 나타나지 않는다. 반면에 코로나가 계속 진행되는 상황 속에서 취임한 Biden의 연설문은 positive keyword에 virus가 포함되어 있으므로 코로나로 인해 바이러스 관련해서 많이 언급했음을 알 수 있다.

comparison cloud에서 주의해야 할 점은 앞에서 말했듯이 배타적인 성질을 가지고 있기 때문에 Trump와 Biden 둘 다 어떤 특정 단어를 많이 사용했더라도 더 많이 사용한 한 사람에게만 positive keyword로 출력된다는 점이다. 따라서 위에서 볼 수 있듯이 Trump에서만 america, american이 positive keyword로 나온다고 해서 Biden이 america, american이라는 어휘를 적게 쓰게 절대 아니라는 것을 주의하고 comparison cloud를 해석해야 한다.

III. 결론

Trump와 Biden의 대통령 취임연설문을 4가지 방법을 거쳐 분석해보았다. 첫 번째로는 불용어를 제거하여 빈도분석을 함으로써 고빈도에 공통적으로 나타나는 contents word는 america, american, americans임을 알 수 있었다. america 관련 단어가 고빈도로 나타나는 이유는 미국인들을 대상으로 미국에 대한 정책이나 미국인들에게 하고자 하는 이야기가 많이 담겨져 있기 때문이고 볼 수 있다.

두 번째로는 bi-gram을 통해 앞뒤에 연달아 연쇄된 두 어휘들을 wordcloud를 이용해 확인해보았다. 여기서 주목할 만한 사실은 Trump는 we will이라는 표현을 많이 사용하여 국민들과 함께 해 나갈 것이라는 확고한 의지를 표현했다고 유추할 수 있다. Biden은 we can이라는 표현을 더 많이 사용하여 국민들에게 할 수 있다는 용기를 북돋아주는 표현을 사용했다고 볼 수 있다. 또한 Trump와 Biden 둘 다 america관련 어휘가 포함된 bi-gram이 많이 나타나는 것을 확인할 수 있다.

세 번째로는 america관련 어휘를 중심으로 설정해 언어분석을 하면 Trump와 Biden이 추진하고자 하는 미국정책이나 미국인들에게 하고 싶은 이야기를 더 명확히 관찰할 수 있을 것이라 생각했다. Trump는 중심어 근처에서 workers, make, first에 해당하는 MI-score이 높게 나왔다. 이는 트럼프가 미국의 노동자나 실업자를 주요대상으로 여기고 있고 미국 우선주의를 토대로 미국을 어떤 방향으로 만들어 나갈지에 대해 이야기하는 것으로 유추해볼 수 있다. Biden은 중심어 근처에서 fellow, story에 해당하는 MI-score이 높게 나왔다. Biden은 국민들을 지칭할 때 fellow라는 어휘를 많이 사용했고 미국의 역사에 대한 이야기를 했음을 확인할 수 있다.

네 번째로는 comparison cloud를 통해 키워드 분석을 했을 때 Trump의 positive keyword로 will, protected, jobs, workers를 찾아볼 수 있다. 앞에서 말했듯이 Trump는 will을 통해 확고한 의지를 나타냈으며 미국을 보호하는 정책을 펼쳤고 근로자와 실업자와, 이와 관련된 일자리에 대해 이야기를 많이 했음을 알 수 있다. Biden의 positive keyword에는 can, unity, uniting, democracy, virus를

찾아볼 수 있다. Biden은 can을 통해 할 수 있다는 능력을 나타냈으며 용기를 복돋아주었고 민주주의와 통합을 매우 강조했음을 엿볼 수 있다. 또한 현재 코로나가 계속 진행되는 상황이 반영되었다는 것 또한 확인할 수 있다.

이처럼 Trump와 Biden의 취임연설문을 비교함으로써 각자 강조하는 내용이 무엇인지, 어떤 정책을 추구하는지, we will과 we can의 사용빈도 차이, 공통으로 많이 사용한 어휘가 무엇인지 등을 발견할 수 있었다. 미국 대통령의 취임연설문을 직접 가공하고 분석하는 과정을 통해 궁금했던 점들을 하나씩 해결해나갈 수 있어서 매우 값진 경험이었다.