토픽 분석을 활용한 대선 토론회 평가

2015580012 통계학과 박수영
2015120039 행정학과 변주연
2015280079 경제학부 정지원
2015920061 컴퓨터과학부 현아연

목차

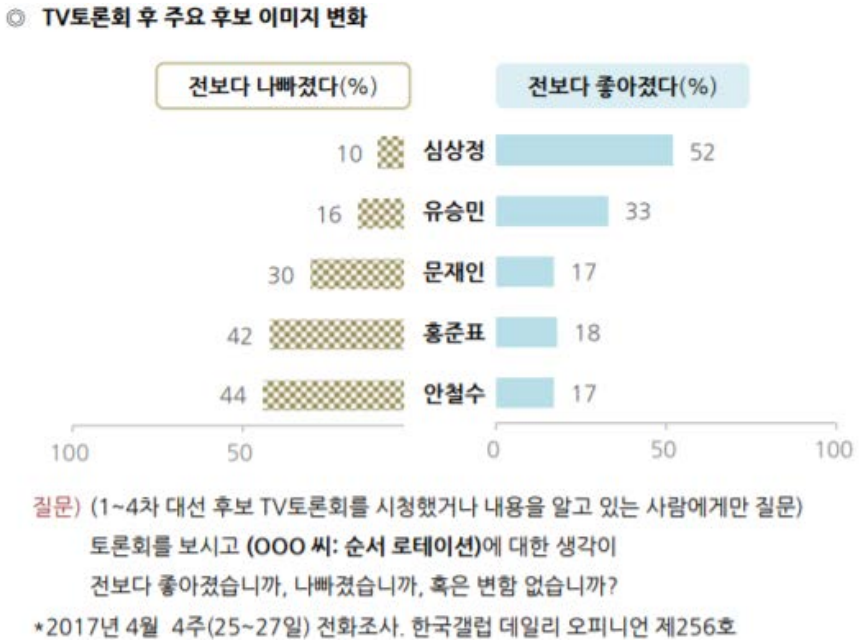
Ⅰ. 서론 1
Ⅱ. 본론 3
1. 데이터 및 변수 설명 3
2. EDA 3
2.1 변수별 탐색 – 전체적 3
2.2 후보별 탐색 6
2.3 단어들의 단순 빈도 분석 17
2.4 연관 단어 네트워크 분석 18
3. 회차별, 후보별 토픽 분석(LDA/CTM) 19
3.1 전 회차 토론 토픽 분석 19
3.2 회차별 토픽분석 21
3.3. 후보별 토픽분석 22
3.4 후보 별 토픽 네트워크 분석 25
4. 추가 분석 26
4.1네거티브 토픽 분석 26
4.2 문재인 후보가 받은 질문 토픽 분석 27
Ⅲ. 결론 29

Ⅰ. 서론

헌정 사상 최초로 대통령이 임기 중 탄핵 소추로 인하여 조기에 시행된 2017년 대한민국 제 19대 대통령 선거는 많은 국민들의 관심 속에서 치러졌다. 많은 시민 참여를 이끌어내며 20년 만에 투표율 77.2%라는 최고기록을 세웠으며, 그 과정에서 후보들의 대선토론회도 주요 이슈로 떠올랐다. 19대 대선은 유권자들이 이전의 대선에 비해 후보자들에 대해서 알아가는 기간이 짧았다. 그렇기 때문에 제약된 시간에 후보에 대해 비교적 많은 정보를

얻을 수 있는 대선토론회에 유권자들의 많은 관심이 모아졌다. 이전 대선까지는 토론회가 후보자들의 지지율에 변화에 큰 영향력을 미치지 않는다는 것이 선거전문가들의 의견이었다. 그러나, 이례적으로 19대 대선의 토론회는 매 회차를 거듭하면서 후보자들의 지지율에 가시적인 변화를 이끌어냈다.

이에 본 분석에서는 19대 대선 토론회 1차, 2차, 3차, 4차 데이터를 활용하여 이에 대한 유의미한 결과를 도출해 보고자한다. 한국 갤럽조사연구소에서 4월 25부터 27일에 걸쳐서 전국의 1006명을 대상으로 조사 실시한 대선 후보들의 TV 토론회 시청에 대한 조사에 따르면 '거의 다 시청했다'가 41%, '일부 시청 했다'가 43%로 나타났다. 이를 바탕으로 국민들의 대선토론회에 대한 관심도가 매우 높음을 알 수 있다. 또한 <그림 1>과 같이 대선토론회에 따라 유권자들의 주요 후보의 이미지 인식이 변화했음을 알 수 있다.



[그림 1]

이로부터 대선토론회가 유권자들의 선택에 유의미한 영향을 미친다는 것을 추론해볼 수 있다. 따라서 본 분석에서는 네 차례에 걸친 19대 대선토론회 데이터를 활용하여 후보자 지지율의 변화를 이끈 대선토론회의 요인들을 살펴보는 것을 연구 목적으로 한다. 대선토론회의 영향력을 판단하는 척도는 한국갤럽에서 실시한 후보자 지지율 조사 결과 [그림 2]를 참고하였다.

	문재인	홍준표	안철수	유승민	심상정
04월 11일	40.60%	8.70%	34.40%	2.90%	3.50%
1차후(4/14)	40%	7%	37%	3%	3%
2차후(4/21)	41%	9%	30%	3%	4%
3차중후(4/24)	39%	14%	31%	4%	3%
4차후(4/28)	40%	12%	24%	4%	7%

[그림 2]

[그림 2]에 따르면, 대선 토론회가 실시되기 이전인 4월 11일 여론조사 결과에서 문재인 후보자가 가장 높은 지지율을 보였고, 유승민 후보가 가장 낮은 지지율을 보였다.

위 여론조사 결과를 바탕으로, 대선후보자 토론회가 토론회 직후의 지지율에 영향을 미치었다고 볼 수 있다. 토론회에서 어떠한 내용이나 태도가 각 후보자에게 긍정적이거나 부정적인 지지율의 변화를 일으켰는지를 파악하는 것을 본 보고서의 목적으로 한다.

Ⅱ. 본론

1. 데이터 및 변수 설명

본 연구에서는 19대 대선 TV토론회 1차, 2차, 3차, 4차의 토론내용을 기록한 데이터를 사용하였다. 주 변수와 도메인은 아래의 표와 같다.

변수명	도메인
From	사회자, 문재인, 심상정, 안철수, 유승민, 홍준표
To	국민, 문재인, 심상정, 안철수, 유승민, 홍준표
Section	negative, 경제, 공약, 교육, 노동, 마무리, 안보, 정치, 공약, 전체
Status	질문, 답변, 공약, 공통질문, 마무리

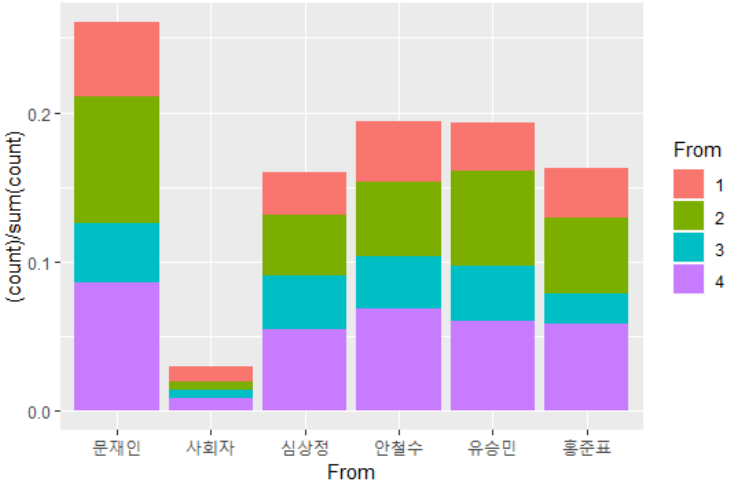
From은 발언자를 의미하며, To는 발언의 대상이다. Section은 논제를 의미하나, 논외 주제인 경우 negative로 표현하였다. Status는 발언의 목적을 의미한다.

자료 분석을 위해 데이터를 분석하기 손쉽도록 변환하였다. 첫째, 1회차에서 negative 변수가 따로 나타나는 것에 대하여 section으로 옮긴 후 변수를 제거하였고, 둘째, 이미 알려진 날짜 변수인 Date 열을 제거하였다. 셋째, 데이터 불러오기 단계에서 생성된 알 수 없는 NULL 행과 열을 제거하였고, 넷째, To가 negative로 오기입된 행을 제거하는 작업을 거쳤다.

2. EDA

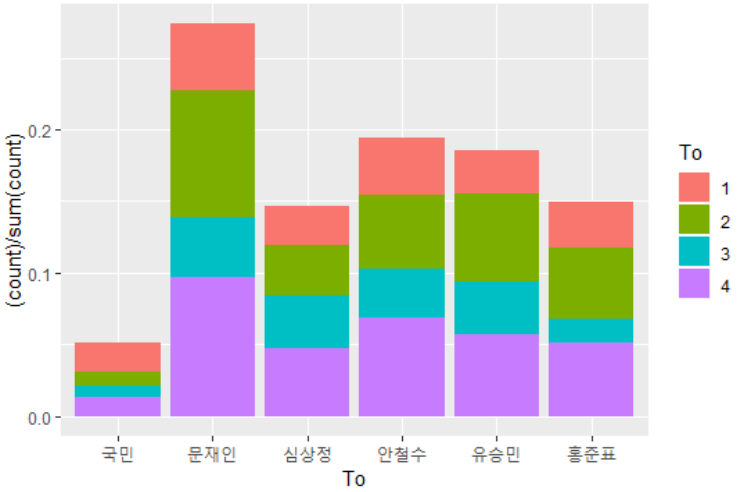
2.1 변수별 탐색 – 전체적인 경향 파악

2.1.1 From 변수



위의 그래프는 From 변수의 상대빈도를 막대그래프로 나타낸 것이다. 전체적인 막대의 크기는 전 회차 토론회의 각 후보별 발언 상대빈도를 의미한다. 후보 별로 볼 때, 문재인 후보가 가장 많이 질문을 하거나 답변을 한 주체였음을을 알 수 있으며, 뒤를 이어 안철수, 유승민 후보의 발언이 많았다는 것을 확인해 볼 수 있다. 회차별로 볼 때, 후보 간 가장 발언의 양에 차이가 많이 나타난 회차는 초록색으로 표시된 2회차임을 알 수 있다. 문재인 청문회가 아니냐는 이야기가 나올 만큼 후보자들의 많은 공격적 질문이 집중되었던 상황에서 여러 다른 후보들의 공격적 질문에 대한 문재인 후보의 답변 발언이 많았던 것을 확인해 볼 수 있다.

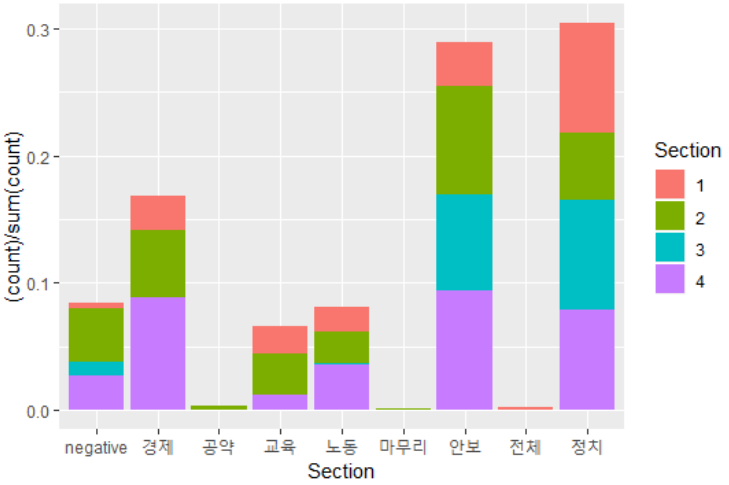
2.1.2 To 변수



위의 그래프는 4회에 걸친 토론회에서 각 후보별 질문 받은 대상의 상대빈도와 후보 당 질문을 받은 상대빈도를 보여준다. 꾸준히 높은 지지율을 받아온 문재인 후보가 다른 후보자들의 공통적인 경쟁 대상이 되어 많은 후보자들의 발언이 문재인을 향해 있었음을 시각적으로 확인해볼 수 있다. 앞서 언급한 것과 같이 특히 처음으로 원고없이 시행된 자유토론방식으로 진행된 2차 토론에서는 많은 질문들이 문재인 후보에게 집중되었다고 이해할 수 있다. 한편, 후보별 전체 막대의 크기에 비하여 3회차에서 유독 후보들에게 질문을 받지 못하여 가장 작은 비율의 대상자가 된 홍준표 후보에 관해서는 왕따설이 돌기도 했다.

후보자들이 다른 후보자들이 아닌 국민들에게 호소하는 내용은 발언대상자를 '국민'으로 명명하여 따로 분류하였는데, 첫 사전녹화로 진행된 1차 토론에서는 다른 회차에 비해 자주 등장하였다. 반면, 2차 이후부터는 회차별로 비슷한 비중으로 등장했음을 알 수 있다. 원고와 함께 진행된 1차 이후부터는 각 후보자들간의 자유 토론으로 이루어져 국민에게 자신의 정책 포부를 전달하기보다는 서로를 향한 열띤 토론이 진행되었다고 유추할 수 있다.

2.1.3 Section

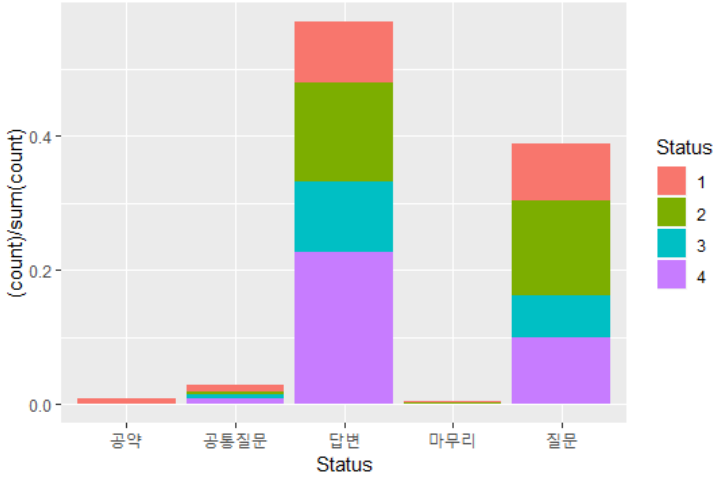


토론회의 내용이 어떠한 하위주제로 세분화되어 이야기되었는지를 section으로 분류하여 회차별로 나타내었다. 먼저 토론의 흐름에 크게 벗어나는 내용을 담은 negative의 경우를 보면, 두드러지게 2회차에서 많이 발언되었으며, 4회차가 그 뒤를 잇는 것을 확인할 수 있다. 1회차의 경우, 원고가 준비된 상태에서 진행되었기 때문에 서로를 비방하거나, 논점을 흐리는 이야기를 하는 경우가 적게 나타났음을 알 수 있다. 한편, 본격적인 negative발언은 2회차에서 정점을 찍는다. 2회차의 상대빈도를 나타내는 초록색 부분을 포함하는 경제, 교육, 노동 등에서의 비중과 비교해 볼 때 negative발언의 비중이 뒤지지 않는 것을 볼 수 있다. 본격 적인 정책논쟁에 대한 토론보다는 후보간 서로에 대한 공격에 치중된 토론이었음을 알 수 있다. 위의 그래프 상으로 negative 발언에 대한 자세한 내용은 알 수 없으나, 문재인 청문회에 준하는 발언이 이어지는 과정에서 빈번하게 발생한 것은 아닐까 하고 추측해 볼 수 있다.

negative를 제외하고는 정치, 안보, 경제, 노동, 교육의 순으로 많이 논의되었다. 1회차에서는 정치가 가장 많이 논의되었다. 2회차에서는 1회차와 비교하여 안보에 대해 활발히 논의가 이루어 진 것을 알 수있다. 북한주적, 핵실험, 사드관련 주제로 후보자 간에 많은 문답이 오고 갔던 것이 원인이라고 생각한다. 2회차에서는 안보의 뒤를 이어 정치 경제, 교육, 노동 등 다양한 주제가 빠짐없이 논의된 것이 확인된다. 1회차와 달리 질의 응답의 순서 없이 이뤄진 자유토론으로 다양한 주제와 의견이 오고 간 것에서 그 이유를 찾을 수 있을 것이다. 3회차 토론에서는 토론의 시작과 동시에 홍준표 후보의 사퇴를 요구하는 발언, 갑철수와 MB아바타 등 negative적 발언이 이루어지기도 했으나, 2차에서는 자유토론형식이 아니고 논의 주제가 외교안보 현안과 정치개혁 현안으로 정의된 상태에서 진행되었기 때문에 본 주제에 적합하게 안보와 정치분야에서의 토론이 활발하게 진행된 것을 확인할 수 있었다. 토론이 가장 토론다웠다고 평가받는 4차 토론의 경우 정치, 안보, 경제등이 비슷한 비중으로 적절히 논의되었으며, 노동과 교육분야와 관련하여서도 후보자들이 자신의 의견을 낸것을 확인해볼 수 있다. 3회차에 비하여 4회차에서는 다른 후보자의 이미지에 타격을

주는 negative 발언들이 이어졌지만, 서로의 공약을 검증하는데 의의가 있었다는 토론으로 평가받는다.

2.1.4 Status



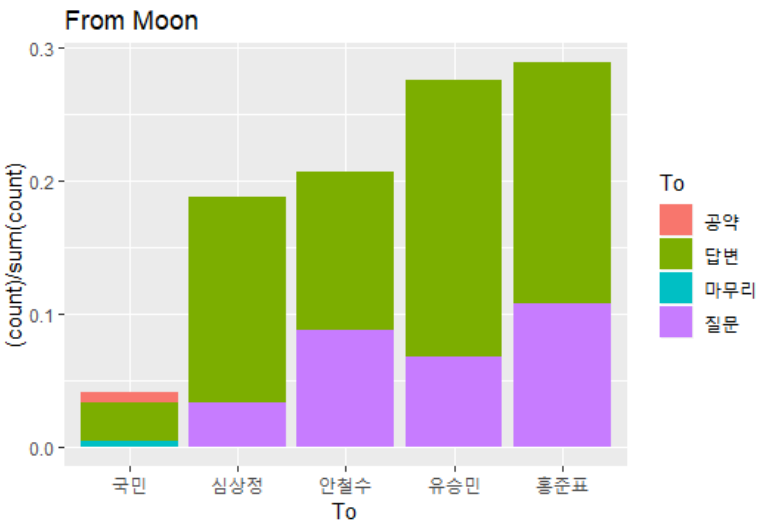
Status는 발언이 공약, 공통질문, 답변, 마무리, 질문 중 어디에 속했는지를 모여주는 것으로 회차별 Status의 빈도를 살펴볼 수 있다. '공약'이 두드러지게 언급된 회차는 1회차임을 확인해 볼 수 있다. 이는 1회차 토론에서 진행된 정책검증 토론과정이 있었기 때문이다. 자신의 공약을 3분간 발표하고, 나머지 후보들의 질문으로 이어져 공약에 대한 검증이 이뤄졌기 때문이라고 생각할 수 있다. 공약으로 분류되는 발언은 원고와 함께 진행된 토론의 특성상 1차 토론에서 후보자들이 자신의 공약을 국민들에게 알리기 위해 많이 언급된 것으로 보여진다. 각 회차를 모두 합친 막대그래프의 크기를 볼 때 '질문'보다는 질문에 대한 답변과 반박 또는 동의로 이루어진 '답변'이 많아 하나의 질문에 대하여 후보간의 치열한 공방이 펼쳐졌음을 알 수 있다. 각 회차별로 질문과 답변의 비중을 비교해볼 수 있다. 1회차의 경우 질문과 답변이 비슷한 분량으로 채워졌다. 정책검증토론이 질문 1분30초, 답변 1분 30초로 질문 답변에 대한 제한된 시간이 있었다는 것이 기여한 결과라고 생각된다. 2회차의 경우도 자유토론으로 진행됐지만, 질문과 답변이 비슷한 비중을 차지했음을 알 수 있다. 3회차와 4회차의 토론의 경우 질문에 비해 답변의 비중이 2배 가까이 크게 나타난 것을 알 수 있다. 승패와 패자가 뚜렷해지는 방향으로 진행된 토론이었던 만큼 여러 질문을 하는 것보다는 한 질문에 대한 쌍방 간의 대화가 깊어질 수 있었다고 추론을 내릴 수 있다.

2.2 후보별 탐색

2.2.1 전 회차 토론

(1) 문재인

1) 질문대상(To) 별 Status(질문/답변)

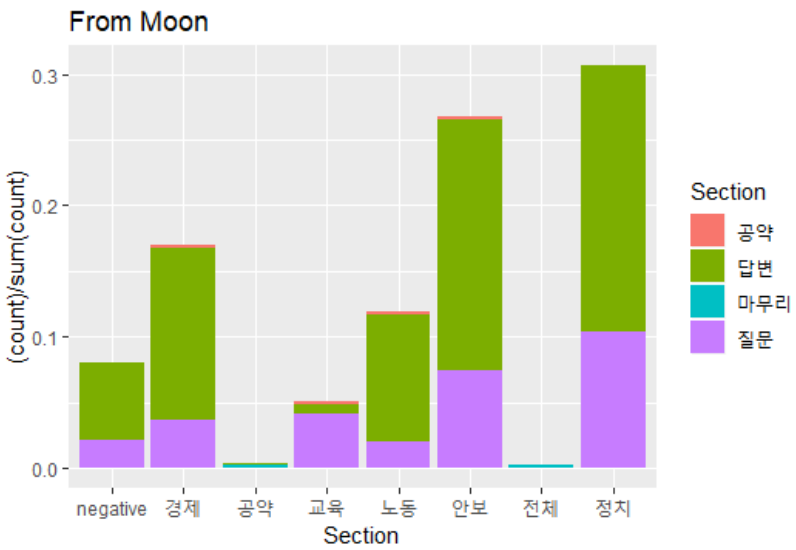


앞서 확인했던 바와 같이 문재인 후보는 발언의 비중과 발언 대상이 되는 비중 모두가 높았다. 이에 구체적으로 문재인 후보가 어떤 대상에게 발언한 비중이 컸는지를 위의 그래프에서 살펴보고자 한다. 문재인 후보는 홍준표 후보, 유승민 후보를 대상으로 하는 발언 비중이 컸으며, 다음으로는 안철수 후보와 심상정 후보에 대한 비슷한 발언 비중을 보였다. 두드러지게 보이는 것은 어떠한 후보를 대상으로 발언을 했던지 전체적으로는 현저하게 질문보다는 답변이 많은 비중을 차지하고 있다는 것이다. 안철수 후보에게는 질문과 답변이 비슷한 비중을 차지하고 있다는 점도 주목해볼 만하다. 국민들의 기대와 달리 토론의 흐름을 이해 못 한 채 질문을 한다는 비판을 받았던 안철수 후보의 경우를 생각해본다면, 토론에 활발하게 참여하지 못하여 문 후보에게 공격적인 질문을 많이 하지 못하였고, 문 후보 역시 안 후보에 대해서는 질문이 기회가 더 많았던 것은 아닐까라고 생각해볼 수 있다.

이에 반해 홍준표 후보나 유승민 후보에게는 각 후보에게 받은 질문을 답하는 데에 토론의 대부분을 할애했다는 것을 파악할 수 있다. 문재인 청문회라고도 불리는 2차 토론을 경험한 후 3차 토론 때 부터는 아예 질문을 모아서 답변하거나, 끊임없는 질문에 대해서는 발언권을 얻고 질문할 것을 요구하는 등의 전략을 취했으나 위의 그래프를 볼때, 다른 후보를 향한 질문보다는 후보들에 질문에 답변을 하는 비중이 더 컸음을 알 수 있었다. 홍준표 후보의 경우 1차 토론때부터 문재인 후보를 겨냥하여 '거짓말하네'의 표현을 비롯한 많은 공격이 이어졌는데, 문재인 후보 역시 이에 응하여 많은 질문을 홍준표 후보에게 던진 것을 확인해 볼 수 있다. 문재인 후보의 입장에서는 상대적으로 저조한 지지율을 보였던 심상정 후보와 유승민 후보가 있었지만, 같은 진보의 입장에서 흑장미 역할을 수행했던 심상정에게 질문은 적게, 반대로 보수의 입장에 섰던 유승민 후보에게는 질문을 보다 많이 했던 것을 확인해 볼 수도 있다.

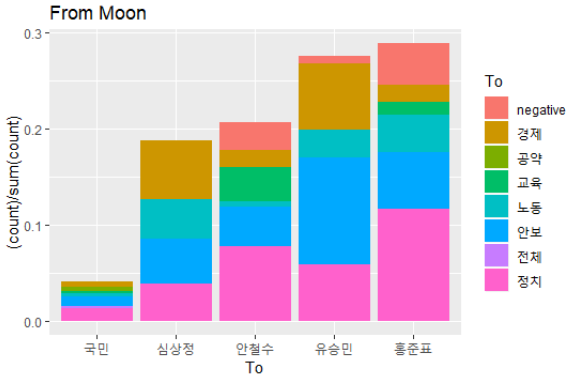
또 하나 눈에 띄는 것은 문재인 후보가 국민에게 답변한 내용이 상당했는데 이는 1차토론 때 한국기자협회소속 언론사들의 질문들을 수합하고 이를 답변하는 과정에서 국민들에게 응답한 것을 국민에게 답변한 것으로 분류된 내용이기 때문이다.

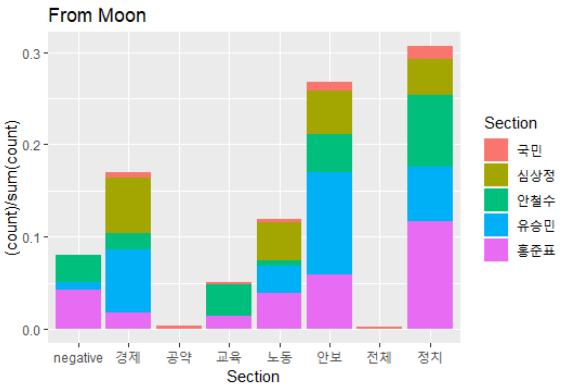
2) 분야 별 Status(질문/답변)



문재인 후보의 발언에서 분야별 비중을 확인해 볼 수 있다. 다른 후보들과 같이 정치와 안보부분에 대하여 발언횟수가 많았지만, 다른 후보들과 약간의 차이는 안보와 정치 모두 비슷한 비중으로 높게 나타났다는 것이다. 뒤를 이어 경제 부문에 대한 발언이 많았으며, 노동분야와 교육분야에 대해서도 빈번하게 언급하였다. 이를 답변비중과 질문비중으로 나누어 볼 수 있다. 질문과 관련하여서는 다음의 그래프에서 살펴보고, 답변한 비중을 주목해서 본다면, 다른 후보들이 문후보에 대해서 정치와 안보, 경제관련 질문에 힘을 실었음을 이해해 볼 수 있다. 또한, 교육은 답변보다는 질문비중이 많은데 이는 문후보가 내세운 교육부분의 공약에 대하여 다른 후보의 공격이 적었다는 것으로 이해해볼 수 있다. 한편, 문 후보의 교육관련 질문이 많았다는 것은 다른 후보의 교육관련 공약에 많은 의문점을 던진 것으로 이해해 볼 수 있다. 한편, negative에 대해서는 토론의 논점과 벗어나는 질문을 하기보다는 답변을 많이 한 것을 알 수 있다. 다른 후보들의 공격성 질문에 negative적 답변을 많이 한 것을 확인해 볼 수 있다. 이는 받은 질문이 negative의 성격을 띄기 때문에 negative적 답변을 했을 수도 있고, 질문이 negative의 성격을 띄지 않았음에도 불구하고 negative적 답변을 했을 가능성도 존재하기 때문에, 이 결과만으로 negative답변에 대한 판단은 유보하기로 한다.

3) 질문대상(To) 별 분야(Section)



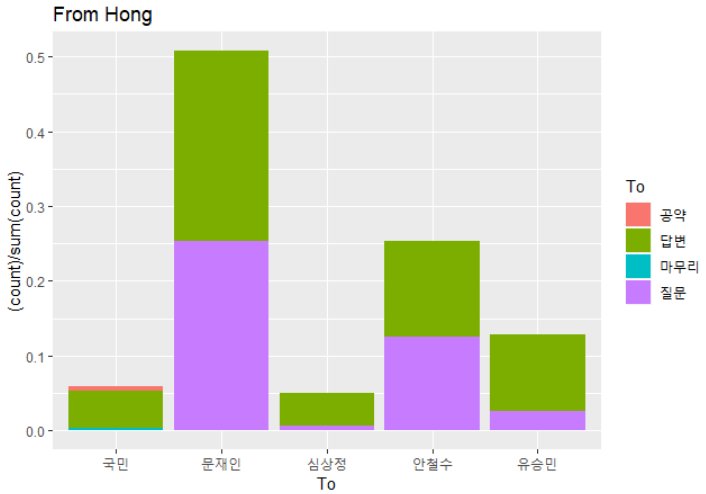


문재인 후보의 질문 대상별 분야에 대한 분석도 가능하다. 문후보의 경우 홍준표 후보에게는 정치와 관련된 발언을 제일 많이 한 것으로 조사되었다. 유승민 후보에게는 안보와 경제관련 질문을 안철수 후보에게는 정치 관련 질문과 다른후보들과는 차별되게 교육 관련 이야기를 많이 묻고 답을 얻었다. 심상정 후보에게는 다른 후보와는 다르게 경제, 노동, 안보, 정치를 포함하며 발언이 분야 별로 비교적 고르게 이루어졌음을 알 수 있다 .

홍준표 후보는 문후보에게 '동성애', '사형제도', '노무현 전 대통령'등 과 관련된 정치적 물음을 많이 던졌는데, 이와 관련된 답변을 비롯한 여러 정치적 발언이 이루어진 것이라고 이해할 수 있다. 유승민 후보는 안보 관련 주제에서 '북한 주적'과 같은 발언으로 토론회에서 논란이 되었는데' 이로 인하여 보안 관련 질의응답을 많이 받았고, 문후보도 이와 관련하여 질문을 많이 한 것으로 추측된다. 심상정 후보와 문재인 후보를 제외한 다른 후보들은 유승민 후보에게 안보관련 발언을 적게 한것을 볼때, 이는 진보 진영과 보수 진영 간의 안보관의 차이가 크게 나타났기 때문이라고 추측할 수 있다. 문재인 후보가 교육 부문과 관련하여 발언한 것의 대부분은 안철수 후보에게 향했음을 밑에의 그래프를 통해 이해해 볼 수 있다. 유승민 후보와 심상정 후보 역시 교육에 관하여 안철수 후보에게 발언이 집중된 것으로 보아 안철수 후보의 교육 관련 정책은 다른 후보들 모두의 공격 대상이 되었음을 알 수 있다. 실제 안철수 후보는 교육 정책을 언급하며 국공립 유치원 설립 자제를 발언하였는데, 많은 젊은 부부들이 안 후보의 이 정책에 반대하였음을 알 수 있다.

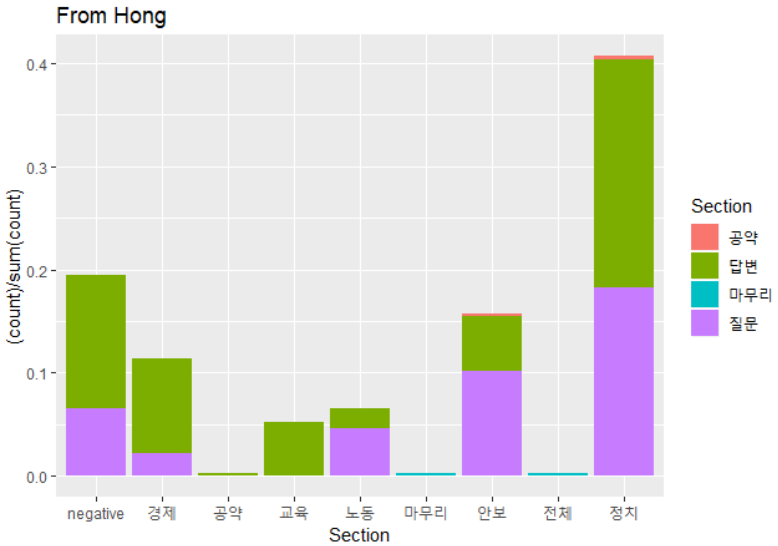
(2) 홍준표

1) 질문대상(To) 별 Status(질문/답변)



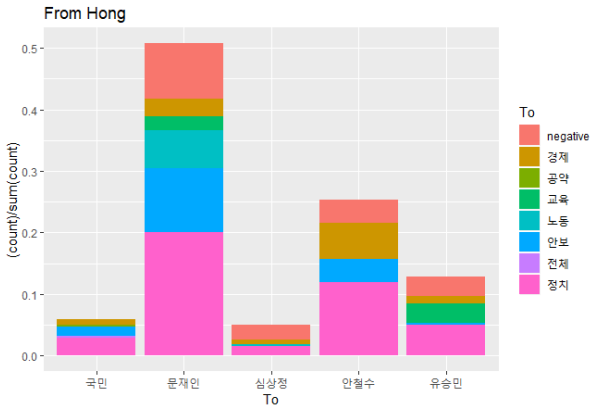
홍준표 후보자의 질문 대상별 분류를 보면 대부분의 질문과 답변이 문재인 후보에게 쏠려있으며, 심 후보나 유 후보에게는 거의 질문조차 하지 않았음을 볼 수 있다. 1차 토론부터 문 후보를 저격하는 질문을 여러 차례 던졌으며, 안철수 후보에게도 질문을 많이 던졌음을 알수 있다. 전체적인 답변과 질문의 비율을볼 때, 위의 문 후보가 질문보다는 전체적으로 답변의 비중이 컸던 것과는 대조되게 질문과 답변이 절반 정도의 비중을 차지했음을 알 수 있다. 이를 통해 볼 때, 홍 후보는 토론에 있어서 상대 후보들에게 공격성 질문을 많이 받기 보다는 자신이 그러한 공격적 질문을 많이 하는 역할을 수행했음을 알 수 있다. 대선토론회에서 자신의 공약을 관철시키기보다는 상대의 정책 및 공략적 허점을 질문하는 방식으로 자신의 입지를 세우고자 하는 태도로 임했음을 위의 그래프를 통해 추론해 볼 수 있다.

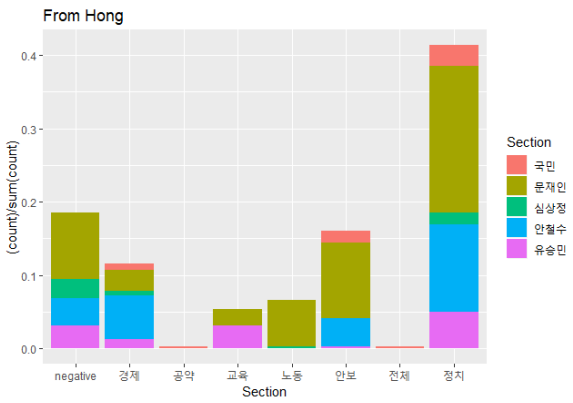
2) 분야 별 Status(질문/답변)



홍준표 후보가 다른 후보들과 두드러지게 차이를 보이는 점은 자신의 발언의 20%가 negative적 발언이었다는 것이다. negative 발언의 최고점을 찍은 '돼지발정제'를 비롯한 엄청난 발언을 했지만, 이를 본 국민들의 반응은 홍 후보가 쌓아왔던 이전까지의 이미지와 크게 차이를 느끼지 않아 별로 놀랍지 않다는 것이었다. 이와 관련해서는 후에 더 살펴볼것이다. 낮술한 시골노인이 아니냐는 비판까지 있었지만, negative적 발언은 멈추지 않았고, 그의 발언의 대부분을 차지했음을 알 수 있다. 그 외에는 정치 관련 발언을 많이 했고, 그 뒤를 이어 안보와 경제, 노동 관련 발언이 차지했다. 각 분야별 발언에서 주목할 점은 역시나 답변보다는 질문이 더 많이 이뤄졌다는 것이다. 위에서 언급한 바와 같이 토론에 있어 상대 후보에게 질문을 많이 던지는 태도를 보였음을 알 수 있다.

3) 질문대상(To) 별 분야(Section)



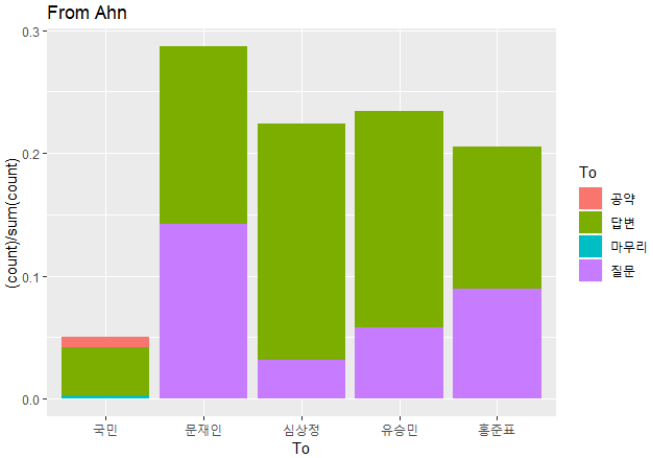


홍준표 후보자를 볼 때 반드시 함께 논의되어야 하는 negative의 비중을 보면, 문재인 후보에게 쏠렸으며, 다른 후보자들에게는 골고루 배분하여 negative적 발언을 했음을 알 수 있다. negative적 발언에 대한 구체적인 내용은 후에 살펴볼 것이기 때문에 여기서는 구체적인 내용설명을 생략한다. 아래의 표를 보면, 대부분의 분야에 대하여 문재인 후보에게 발언한 것이 대부분을 차지함을 알 수 있다. 다음 비중으로 안철수 후보를 대상으로

발언하였는데 그 격차가 상당하다. 홍 후보가 문재인 후보에게 발언한 것을 분야별로 보면 경제, 교육, 노동, 안보, 정치까지 모든 분야에 대해 질문하고 답변했음을 알 수 있다. 말그대로 문재인 후보에게 총력을 다한 것이다. 특별한 분야에 한하여 문후보에게 발언한 것이 아니라 문 후보가 내세운 정책의 모든분야에 대하여 질의응답을 받은 것은 문 후보를 집중공격하는 것이 홍 후보 토론의 시작과 끝이었다고 봐도 과언이 아닐 것이다. 문 후보 다음으로 선택한 발언 대상은 당시 지지율 2위였던 안 후보였는데, 문재인 후보가 안 후보와 교육관련분야에 집중적으로 대화를 나눈 것과 달리 홍 후보는 안 후보에게 교육 관련 논의는 하지 않았음을 알 수 있다. 오히려 경제나 정치적 분야에 대하여 더 묻고 답했던 것을 확인해 볼 수 있다. 또한 오히려 유 후보와 교육 관련대화를 나눴다. 재미있는 점은 토론에서 심상정 후보와의 대화는 거의 찾아보기 어려웠다고 볼 수 있다. 국민에게 호소하는 발언보다도 심상정 후보와의 대화가 더 적었던 점으로 볼때, 홍 후보와 심상정 후보는 서로 간의 질의가 자신의 지지도 상승을 이끌어내는데 큰 역할을 하지 않는다고 판단했다는 것을 알 수 있다.

(3) 안철수

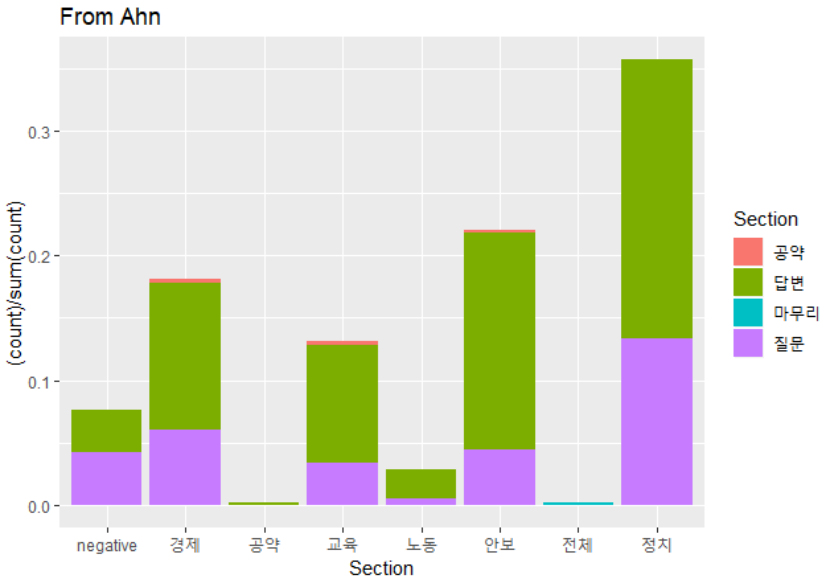
1) 질문대상(To) 별 Status(질문/답변)



위 그래프에서 안철수 후보는 문재인 후보에게 가장 많은 발언을 한 것으로 나타났고, 그 외의 후보들에게는 상대적으로 고르게 발언을 한 것을 알 수 있다. 문 후보에게 질문이 가장 많이 집중된 것으로 보아 안 후보가 문 후보를 가장 유력한 경쟁자로 인식하고 있다는 것을 유추할 수 있다.

문재인 후보에게는 질문 및 답변의 비율이 비교적 고르게 나타난 반면, 심상정 후보 및 유승민 후보에게는 많은 답변을 한 것으로 보아 이는 이 두 후보의 공약과 안 후보의 공약에서 많은 차이가 나타났기 때문이라고 추측할 수 있다. 안후보와 같은 중도보수 또는 보수 진영으로 분류되는 유승민 후보와 홍준표 후보에게도 많은 질문 및 답변이 이루어졌는데, 두 후보가 비슷한 정치적 성향으로 분류되지만, 두 후보와의 공약의 방향성이 달랐기 때문이라고 유추할 수 있다. 이에 대한 더 자세한 추론은 안 후보의 분야별 발언 상대빈도를 통해서 이루어 질 것이다.

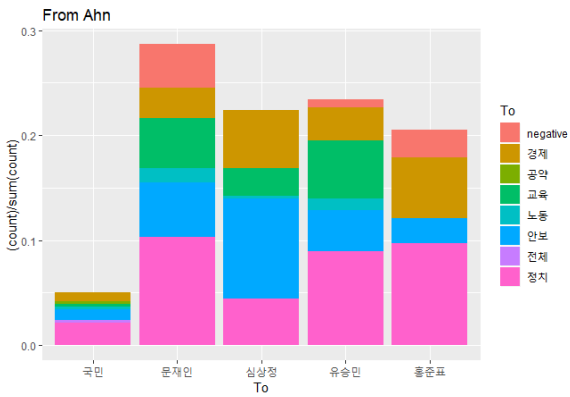
2) 분야 별 Status(질문/답변)

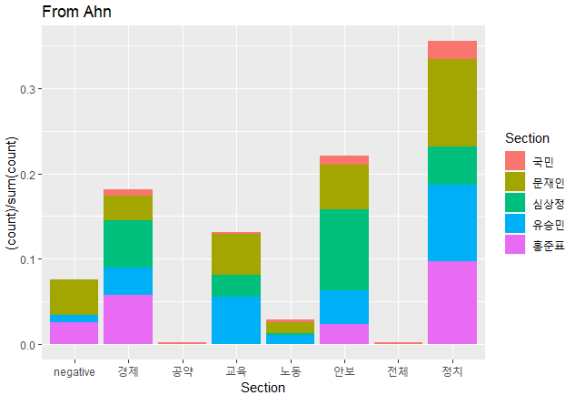


안철수 후보는 정치와 관련된 질문 및 답변이 가장 빈번하게 이루어졌다. 안 후보의 경우에는, 상대적으로 노동 관련 발언보다도 negative 발언이 더 많이 나타났다. 상대적으로 안보에 관한 것은 질문보다는 답변이 많이 이루어졌다는 것도 알 수 있다.

안 후보의 답변 및 질문은 주로 정치에 관한 것으로 이루어져 있다. 토론회 당시 특히 화제가 되었던 “MB 아바타”, “박지원 꼭두각시” 등의 안 후보를 향한 네거티브 발언도 안 후보의 정치입지의 불안정성을 나타내는 것이다. 이처럼 안 후보도 자신의 정치 관련 이슈가 안 후보에게 불리하게 작용할 수 있다고 인식했기 때문에 정치 관련 답변이 다른 분야에 비해 많이 나타났다고 볼 수 있다.

3) 질문대상(To) 별 분야(Section)

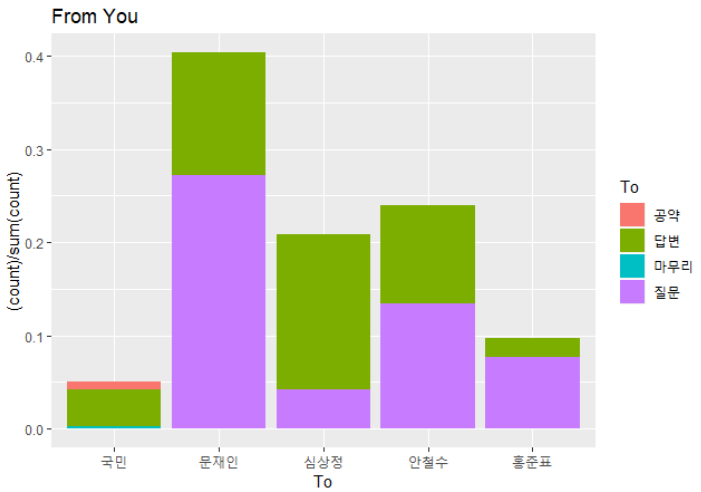




안 후보자는 문재인 후보에게 가장 많이 질문을 하였으며 주로 정치와 안보에 대하여 토론했다. 또한 심상정 후보에게 유난히 안보에 대하여 많이 질문했음을 그래프에서 볼 수 있다. 또한 홍준표 후보자에게는 안보 관련 질문이 유난히 적으며 경제 관련 질문이 타 후보자들보다 많이 나타났다. 안 후보자는 주로 정치에 관련하여 많이 발언하였으며 그 다음으로는 안보, 경제를 들 수 있다.

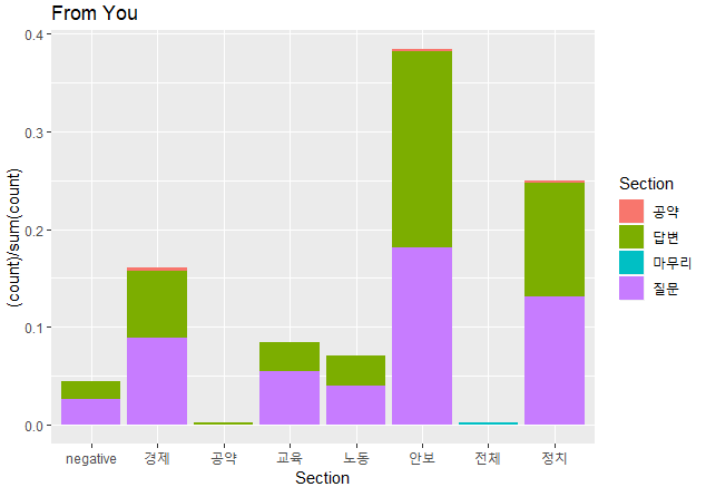
(4) 유승민

1) 질문대상(To) 별 Status(질문/답변)



유승민 후보자는 문재인 후보에게 40%에 달하는 발언을 하였으며 그 중 질문이 약 2/3를 차지한다. 타 후보자 중 심상정 후보자에게 가장 많은 비율의 질문을 받은 것으로 나타난다. 홍준표 후보와 비슷한 모습으로 답변보다는 질문이 더 많은 발언의 비율을 차지했지만, 후보별로 다른 모습을 보인다. 문 후보와 안 후보에게는 질문을 많이 하면서도 동시에 답변을 많이 했지만, 홍준표 후보에게는 질문이 대부분을 차지했음을 알 수 있다. 홍준표 후보에게 많은 질문을 던졌지만, 이와 반대로 홍준표 후보는 문 후보를 공격하는 것에 집중했기에 유승민후보에게 질문을 많이 던지지 않은 것을 확인해볼 수 있다. 홍후보와는 반대로 심상정 후보는 유승민 후보에게 많은 질문을 던져 유후보는 많은 답변을 한 반면, 유 후보는 심 후보에게 많은 질문을 던지지 않았음을 알 수 있다. 이를 통해 볼 때, 유승민 후보는 전반적으로 상대후보에게 질문을 많이 던지는 공격적 태도를 취했지만, 그 공격의 대상이 문재인과 안철수, 홍준표 후보에게 집중되었음을 알 수 있다.

2) 분야 별 Status(질문/답변)

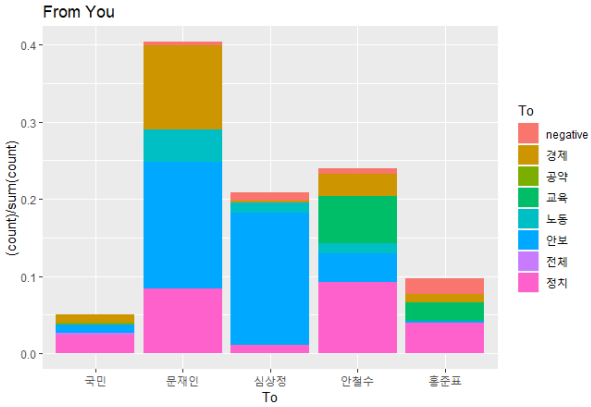


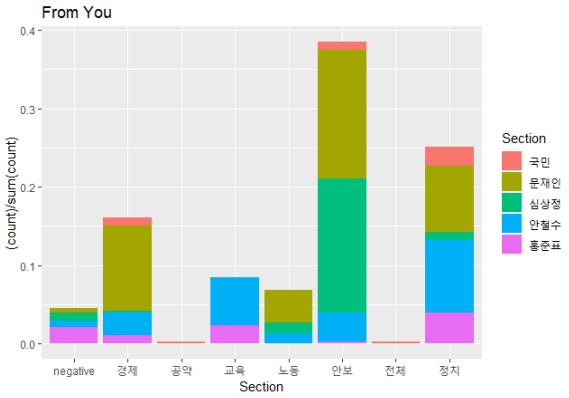
홍준표 후보의 negative발언 비중이 20%였던 것과 대조되게, 유 후보의 전체 발언 대비 negative적 발언은 전체 후보들 중 가장 작은 비율을 차지하였다. 유 후보는 국민들로부터 negative의 발언 비율이 적고, 논제의 적합한 분야 위주로 토론을 진행하여 다른 후보들에 비해 토론의 흐름을 벗어나지 않는 선에서 다양한 분야의 발언을 한 것으로 평가 받아 토론을 잘한다는 평가를 받았다.

유승민 후보자는 안보와 관련한 발언을 가장 많이 하였으며, 그 다음으로는 정치와 경제 발언을 많이 한 것으로 나타났다. 대부분의 분야에 대한 고른 질문과 답변을 한 점도 알 수

있다. 안보와 관련하여 '북한인권결의안', '사드배치'등과 관련하는 당시의 화제되었던 의제 위주로의 구체적인 질문 및 답변이 이루어졌다.

3) 질문대상(To) 별 분야(Section)

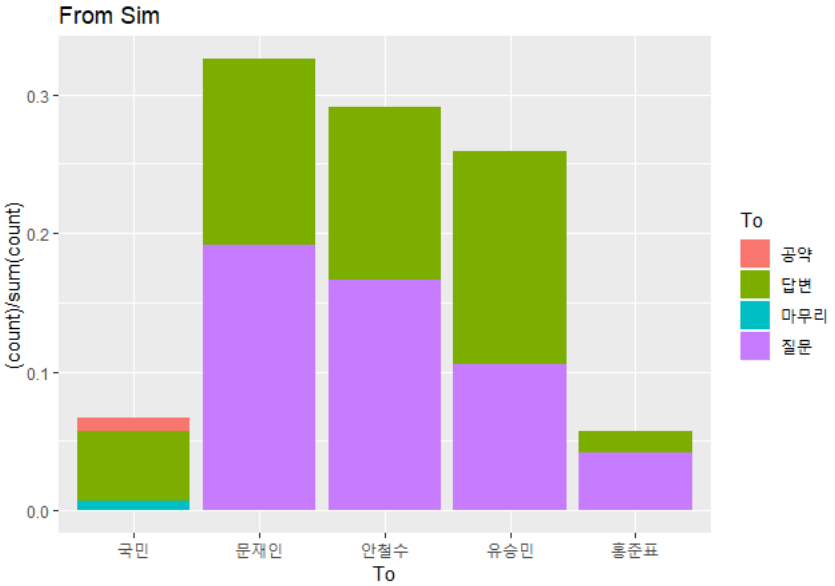




유승민 후보자는 문재인 후보자에게 가장 많은 발언을 하였으며 안보가 주된 내용이었다. 또한 심상정 후보자에게 한 발언 중 거의 대부분을 차지하는 분야도 안보로 나타났다. 안보 정책에서 가장 큰 차이를 보였기 때문에 심상정 의원과의 비중이 가장 크게 나타난 것이다. 또, 심 후보 및 문 후보에 대한 비중이 높은 것으로 보아, 주로 자신과 반대되는 성향의 후보자들에게 많은 질문을 한 것이 나타난다. 또한 각 분야에서는 안보가 가장 높은 비율을 보이며 그 다음이 정치 분야로 나타났다.

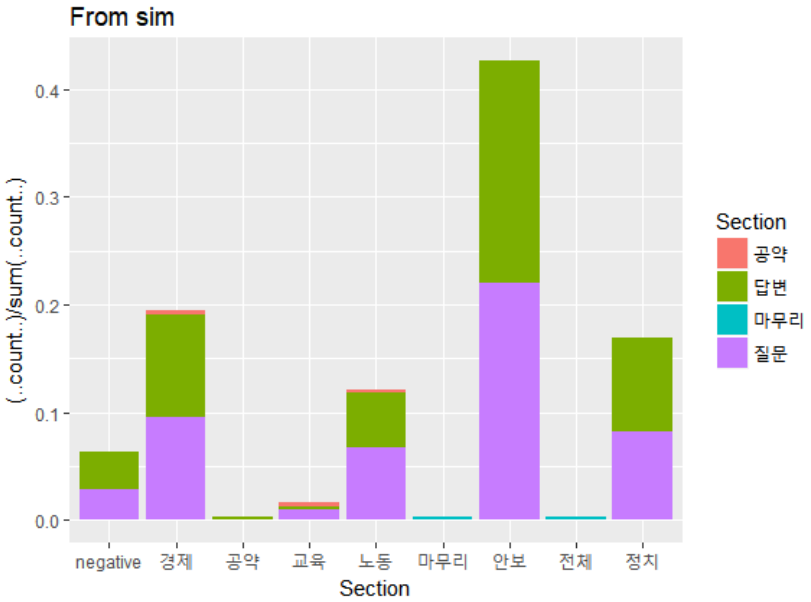
(5) 심상정

1) 질문대상(To) 별 Status(질문/답변)



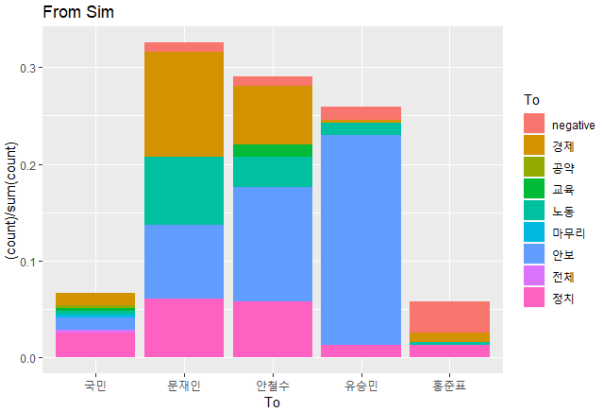
심상정 후보자는 홍준표 후보자를 제외한 세 후보자에게 많은 질문과 답변을 했다. 홍 후보에게는 국민에게 한 발언의 비중보다도 더 적게 발언한 것을 알 수 있다. 타 후보자와 유사하게 문재인 후보로의 질문과 답변이 많았다. 대체적으로, 다른 후보들에게 질문과 답변의 비중이 비슷하게 나타났다. 그러나 유 후보에게는 전반적으로 답변의 비중이 더 우세했다.

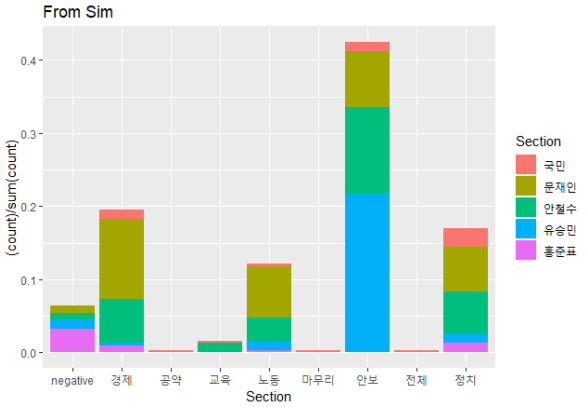
2) 분야 별 Status(질문/답변)



심상정 후보자는 모든 분야에서 질문과 답변을 골고루 한 것을 알 수 있다. 안보 관련 발언 비중이 40%로 매우 높게 나타났는데, 뒤를 이어 경제와 정치 관련 발언이 비슷한 비중을 차지하였다. 반면, 교육에 대한 발언이 비중이 매우 낮게 나타났는데, 특히, 교육 관련 답변이 적은 것으로 보아 후보들로부터 교육 관련 질문을 적게 받았었기 때문이라고 추론할 수 있다.

3) 질문대상(To) 별 분야(Section)

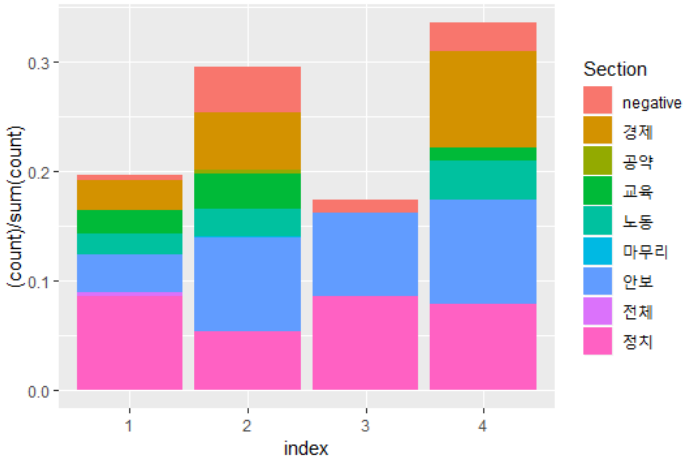




위 그래프에서 눈에 띄게 나타나는 점은 '안보' 발언의 비중이다. 안보 분야에 대해서 중점적으로 발언했으며, 특히 유승민 후보자에게 안보와 관련한 많은 발언을 하였다. 다른 후보들과 비슷한 점으로는 문재인 후보에게 발언의 비중이 가장 높게 나타났다는 것이다.

2.2.2 회차별 탐색

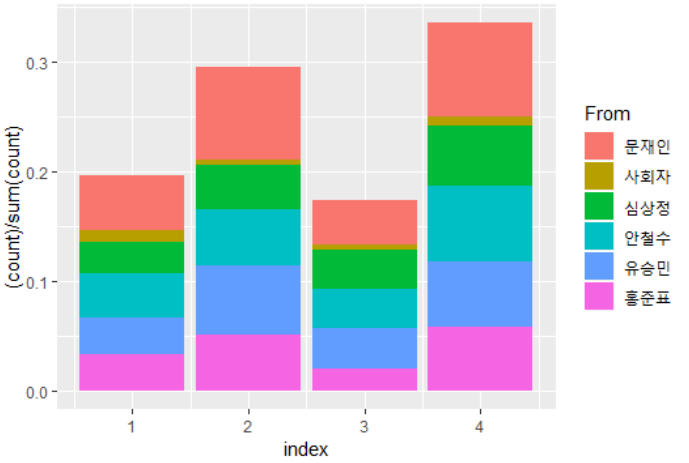
(1)Section



위 그래프는 각 회차마다 후보자들의 발언에 대한 구분, 즉 Section 별로 나타내고 있다. 회차마다 후보자의 발언의 양에 차이가 있었기 때문에, 각 회차에서 막대의 길이가 다르게 나타났다. 총 9가지의 Section 분류 중 마무리는 분류 특성 상, 모든 회차에서 가장 적게 나타났다. 후보들의 공약 중 우선 순위에 해당하는 정치, 안보 및 경제가 상대적으로 높은 빈도로 나타났다.

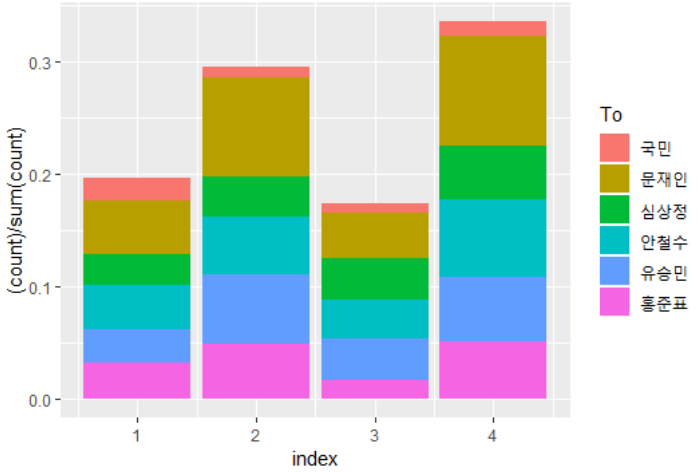
회차에 따라서 상대적으로 가장 많이 언급되는 분야는 다르게 나타났다. 1차 토론에서는 정치가 가장 높은 빈도로 언급되었다. 1차 토론에 비해서 2차와 4차 토론에서는 경제 및 안보의 언급 빈도가 상대적으로 증가한 반면, 정치의 언급 빈도는 감소하였다. 3차 토론의 발언은 마무리를 제외한 9개의 분류 중 negative, 노동, 정치, 안보 분야만 언급되었는데, 특히 정치와 안보 분야로 양분되었다.

(2) From



위 그래프에서는 회차마다 후보자별 발언 횟수를 보여주고 있다. 후보자 별 발언 횟수는 대체적으로 고르게 분포하고 있지만, 모든 회차에서 문재인 후보의 발언 횟수가 가장 크게 나타났다. 그러나 그래프의 축의 값을 고려한다면, 시각적으로 보이는 막대 크기의 차이만큼 실질적인 발언수의 차이가 크지 않다는 것을 알 수 있다.

(3) To



위 그래프는 각 후보가 질문의 대상이 된 횟수를 표현한 그래프이다. 후보자들이 다른 후보자들이 아닌 국민들에게 호소하는 발언은 적게 나타났다. 모든 회차에서 문재인 후보자를 대상으로 한 발언이 가장 많이 나타났다. 2회차를 제외한 다른 회차에서 홍준표 후보가 질문을 받은 빈도 수가 가장 낮게 나타났다. 3회차를 제외한 외차에서 문재인 후보를 제외한 후보자들이 비교적 고르게 질문을 받았다는 것도 알 수 있다.

2.3 단어들의 단순 빈도 분석

가장 많이 나타난 단어들의 단순 빈도 분석을 실시한 후 가시적인 워드클라우드 형태로 나타내었다. 여기서 '우리', '후보님', '말씀', '후보'와 같이 의미가 없고, 상대를 지칭하기 위해 자주 반복되는 단어들이 가장 많이 나타난 것을 직관적으로 알 수 있다. 이러한 단어들은 대선토론회에 대한 토픽 분석이라는 연구목적에 부합하지 않는다. 그러므로 후보자의 특징을 추출하는데에 의미론적으로 크게 관여하지 못하므로 제거하여 분석하도록 연구 방향을 설정하였다. 아래는 이러한 단어를 제거한 뒤 생성한 워드클라우드이다.

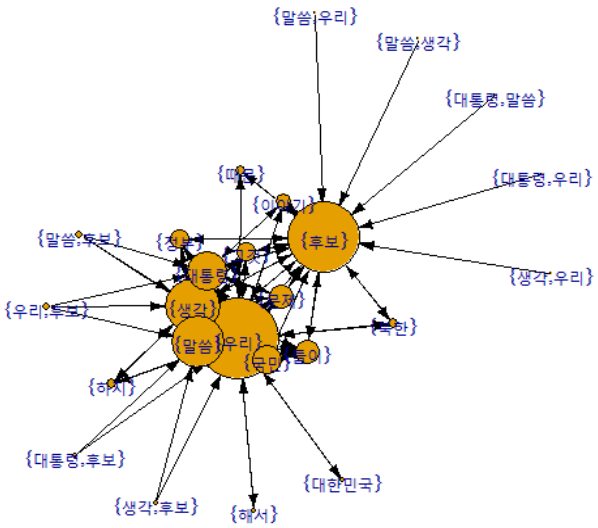
실질적인 의미를 가지지 않기 때문에 불필요하다고 판단한 단어를 제거한 뒤, 워드클라우드를 생성했다. 위 워드클라우드를 통해 일자리 문제나 자국과 북한, 미국 간의 관계 등 후보들의 주요 공약을 간략하게 확인할 수 있다.

```
sort(bigramlist, decreasing=T)[1:20]
        수 있는              여러 가지            그리고 또              그 다음에                  할 수
             55                     44                   41                     32                     31
   대통령이 되면         생각을 합니다    문제에 대해서 어떻게 생각하십니까         그래서 저는
             30                     28                   24                     24                     22
        예 저는            국민 여러분              두 번째            우리 유승민                  한 번
             22                     21                   21                     21                     21
           그 때              수 있다고          이명박 박근혜            이 문제에             그런데 지금
             20                     19                   19                     18                     17
```

개별적인 단어만의 분석에서 의미를 파악할 수 없을 가능성이 있기 때문에, 2개의 단어로 구성된 bigram 중 상위 20개를 찾아냈다. 그 중 '수 있는'과 같은 크게 의미가 없는 단어들도 볼 수 있었으나 전체적으로 의견을 묻는 질문과 관련한 2-gram 단어들이 많이 나왔다고 할 수 있다.

2.4 연관 단어 네트워크 분석

연관성 분석을 통하여 주로 함께 사용하여 연관성이 높다고 판단되는 단어들의 개략적인 네트워크를 파악할 수 있다. 우선 의미없는 단어들을 전처리하기 전에는 워드클라우드 생성 시와 유사하게 '우리', '생각', '말씀'과 같은 말이 중심에 있고 파생되어서 다양한 단어들이 나타난다. 또한 후보와 관련해서는 '후보님 말씀', '우리 후보', '후보님의 생각', '대통령 후보' 와 같은 단어가 연관성이 높게 나타났다.



3. 회차별, 후보별 토픽 분석(LDA/CTM)

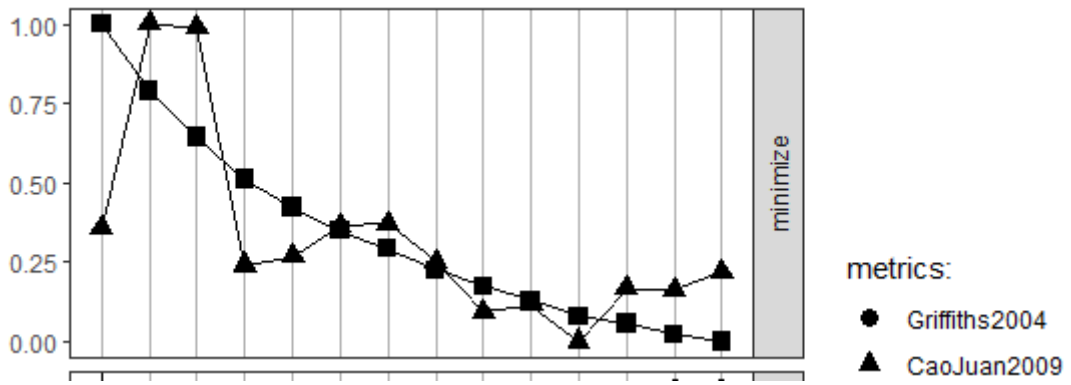
3.1 전 회차 토론 토픽 분석

워드클라우드에 나타나듯, '우리', '후보'와 같이 크게 의미없는 단어들이 나타났다. 따라서 분석에 의미없는 단어들을 제거하여 나타난 유의미한 단어들 중 유의미한 토픽을 추출해내었다.

제거한 단어 목록은 다음과 같다.

그것	때문	그거	하지	말씀	우리	하게	번째	대통령	생각
부분	이야기	하시	그걸	이번	들이	저것	이걸	사실	얘기
입장	질문	이것	다음	확인	후보	대한	민국	그때	거기

토픽 분석 전에 LDA에서 토픽의 개수를 찾기 위해 'ldatuning'이라는 패키지 내장 함수를 이용하였다.



CaoJuan(2009)은 LDA(Latent Dirichlet allocation)에서 최적의 k값을 찾기 위해 베이지안의 비모수적 방법에 기반한 HDP(hierarchical Dirichlet process)와의 혼합비율로

계산한다. 여기서 너무 많은 토픽을 가진다면 단어 간 의미가 서로 겹치게 되므로 식별 가능성을 유지하기에는 너무 많은 상관 관계가 발생할 수 있다. 따라서 k의 수가 너무 적어서 동질성이 떨어질 위험과 반대로 너무 많아서 군집 간 의미가 서로 겹치는 위험을 줄이기 위해, 이 방법을 바탕으로 본 분석에서는 적절한 k값을 3과 5로 설정하였다.

3.1.1 LDA

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
"국민"	"정부"	"북한"	"문재인"	"정부"	"일자리"	"일자리"	"문제"
"개혁"	"노무현"	"미국"	"정책"	"가지"	"공약"	"문제"	"북한"
"교육"	"합의"	"중국"	"안철수"	"중요"	"복지"	"중소기업"	"찬성"
"안철수"	"박근혜"	"핵실험"	"대표"	"국민"	"증세"	"시간"	"문재인"
"사람"	"홍준표"	"외교"	"토론"	"산업혁명"	"문재인"	"경제"	"반대"
"누구"	"문재인"	"문제"	"국민"	"상황"	"박근혜"	"노동"	"안보"
"유승민"	"지원"	"정부"	"전술"	"미래"	"기업"	"비정규직"	"거짓말"
"공약"	"문제"	"안보"	"박지원"	"국가"	"인상"	"성장"	"기권"
"문제"	"이명박"	"검찰"	"정부"	"시대"	"세금"	"임금"	"심상"
"적폐"	"해서"	"배치"	"북한"	"정말"	"국민"	"대기업"	"정의"
"선거"	"정권"	"한반도"	"공약"	"안보"	"부분"	"부분"	"북한인"
"안보"	"잘못"	"해서"	"사건"	"평가"	"노조"	"청년"	"판단"

3.1.2 CTM

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
"국민"	"안보"	"문제"	"일자리"	"정부"
"북한"	"문재인"	"정원"	"문제"	"국민"
"핵실험"	"토론"	"북한"	"정부"	"북한"
"중국"	"개월"	"검찰"	"교육"	"박근혜"
"문제"	"국민"	"국민"	"중소기업"	"반대"
"경제"	"문제"	"평가"	"기업"	"공약"
"개혁"	"대북"	"사건"	"지원"	"정책"
"임금"	"공약"	"동맹"	"개혁"	"안철수"
"증세"	"해서"	"중국"	"국가"	"합의"
"외교"	"군대"	"안보"	"해서"	"유승민"
"공약"	"송금"	"국회"	"거짓말"	"사람"
"복지"	"가지"	"외교"	"산업혁명"	"전술"

상관 토픽 모형(CTM)의 결과표는 위와 같다. LDA를 참고하여 해석하면, 첫번째 토픽은 '국민', '북한', '개혁' 등 전반적인 정치 이슈와 관련되어 있고, 두번째 토픽은 '안보', '문재인', '대북', '군대' 등 문재인 후보의 대북 및 군사 정책 공약과 관련되어 있다. 셋째는 '청원', ' 검찰', '국민', '평가', '안보' 등 안보 및 외교에 대한 주제이며, 다음은 '일자리', '교육', '기업', '지원', '개혁' 등 노동 및 교육 정책과, 마지막 토픽은 '정부', '반대', '공약', '정책' 등 선거 전반과 동시에 '박근혜', '안철수', '유승민' 등의 보수세력과 관련되게 나타났다. 따라서 분석 결과, 전 회차 토론회에서는 크게 전반적 정치 이슈와 대북/군사 정책, 사회복지 및 노동 정책 그리고 선거와 관련한 토론이 중심 주제였다고 볼 수 있다.

3.2 회차별 토픽분석

3.2.1 LDA

1차

Topic 1	Topic 2	Topic 3
"정부"	"국민"	"일자리"
"주장"	"북한"	"국민"
"정책"	"안보"	"성장"
"문제"	"사면권"	"중소기업"
"박근혜"	"언론"	"시간"
"재벌"	"정부"	"교육"
"합의"	"국정"	"인상"
"노무현"	"세력"	"소득"
"개혁"	"적폐"	"기업"
"배치"	"정권"	"복지"
"반대"	"박근혜"	"공약"
"탕감"	"민주"	"경제"

2차

Topic 1	Topic 2	Topic 3
"교육"	"합의"	"문제"
"북한"	"북한"	"중국"
"핵실험"	"반대"	"대북"
"국민"	"일자리"	"정부"
"공약"	"전술"	"안철수"
"외교"	"세금"	"국민"
"정부"	"사람"	"국가보안법"
"가능성"	"무상급식"	"박지원"
"문재인"	"발표"	"송금"
"문제"	"토론"	"중요"
"무엇"	"주적"	"북한"
"산업혁명"	"사회"	"문재인"

3차

"검찰"	"문제"	"정부"
"개혁"	"국민"	"북한"
"문제"	"공약"	"문제"
"사건"	"대표"	"거짓말"
"수사"	"북한"	"가지"
"정원"	"안철수"	"지원"
"문재인"	"안보"	"노무현"
"청와대"	"박지원"	"중요"
"국민"	"정부"	"블랙리스트"
"사람"	"이후"	"과거"
"홍준표"	"핵실험"	"이거"
"국회"	"정치"	"홍준표"

4차

"공약"	"미국"	"일자리"
"일자리"	"안보"	"정부"
"정책"	"문제"	"문제"
"토론"	"사람"	"재벌"
"문제"	"전술"	"경제"
"국민"	"가지"	"퍼센트"
"정부"	"북한"	"투자"
"시간"	"정책"	"기업"
"박근혜"	"동의"	"민간"
"공무원"	"중국"	"반대"
"정도"	"정부"	"노조"
"부담"	"중요"	"해서"

각 회차별 LDA 토픽 분석을 실시한 결과는 위와 같다. 1회차의 분석 결과로 미루어 보아, 기존 정부와 기업에 대한 정책, 고용 정책 및 경제 정책 등과 관련한 주제가 주를 이뤘다고 할 수 있다. 2회차에서는 대북 정책을 포함한 외교 정책과 사회복지 정책이, 3회차에서는 입법부, 사법부, 행정부를 비롯한 정부와 특정 후보 및 공약이 관련한 주제로 나타났고, 마지막 4회차에서는 각 후보들의 정책 공약, 안보와 외교 그리고 노동을 포함한 경제 분야와 관련된 주제가 나타났다.

3.2.2 CTM

1차	2차

Topic 1	Topic 2	Topic 3
"북한"	"안보"	"국민"
"국민"	"문제"	"일자리"
"공약"	"해서"	"정부"
"합의"	"국민"	"중소기업"
"배치"	"정책"	"성장"
"반대"	"개혁"	"사면권"
"복지"	"정부"	"언론"
"비정규직"	"교육"	"박근혜"
"미국"	"시간"	"공영방송"
"대응"	"보수"	"경제"
"필요"	"중요"	"서민"
"증세"	"국가"	"법원"

Topic 1	Topic 2	Topic 3
"북한"	"교육"	"국민"
"핵실험"	"문제"	"합의"
"중국"	"국가보안법"	"정부"
"외교"	"폐지"	"대북"
"가지"	"무상급식"	"시간"
"일자리"	"반대"	"문재인"
"정원"	"누구"	"안보"
"문제"	"문재인"	"안철수"
"미국"	"해결"	"송금"
"발표"	"거짓말"	"박지원"
"하나"	"외교"	"선거"
"정부"	"공약"	"정책"

3차

Topic 1	Topic 2	Topic 3
"정부"	"검찰"	"국민"
"북한"	"개혁"	"문제"
"문제"	"정원"	"북한"
"거짓말"	"안보"	"안철수"
"노무현"	"수사"	"대표"
"사건"	"문제"	"공약"
"지원"	"청와대"	"선거"
"문재인"	"개헌"	"중요"
"안철수"	"경찰"	"정부"
"장관"	"국민"	"중국"
"색깔론"	"권력"	"미국"
"박근혜"	"하나"	"가지"

4차

Topic 1	Topic 2	Topic 3
"일자리"	"문제"	"정부"
"문제"	"재벌"	"미국"
"공약"	"퍼센트"	"정책"
"정부"	"경제"	"전술"
"국민"	"누구"	"북한"
"공무원"	"노무현"	"반대"
"민간"	"개혁"	"박근혜"
"투자"	"동맹"	"그것"
"안보"	"임금"	"사람"
"사람"	"정도"	"임금"
"중요"	"안보"	"해서"
"토론"	"정부"	"토론"

각 회차별 CTM 토픽 분석을 실시하였을 때 결과는 위와 같다. 각 회차에서의 분석 결과로 주제를 미루어 짐작해보았을 때, 1회차에서는 '북한','공약','반대','복지' 등 정책과 관련되며 2회차에서는 대북과 외교, 교육정책과 선거와 후보, 3회차에서는 지난 정권, 국가 안보, 4회차에서는 일자리, 기업 등 경제 정책 및 개혁과 관련한 주제들이 주로 다루어졌다.

따라서 최종적으로 각 회차별로 1회차에서는 기업 개혁과 고용 경제, 2회차에서는 사회복지 정책, 3회차에서는 국가 안보와 안철수 후보, 4회차에서는 일자리 정책 등이 주로 토론되었으며 매 회차 대북 외교 및 안보와 관련한 주제는 꾸준히 등장했다.

3.3. 후보별 토픽분석

3.3.1. LDA

```
> #문재인
> terms(lda.out, 12)
      Topic 1  Topic 2  Topic 3
 [1,] "일자리" "안보"   "성장"
 [2,] "문제"   "국민"   "합의"
 [3,] "정부"   "박근혜" "일자리"
 [4,] "북한"   "정부"   "해서"
 [5,] "퍼센트" "시간"   "국민"
 [6,] "필요"   "정권"   "정부"
 [7,] "중국"   "나라"   "노무현"
 [8,] "핵실험" "해결"   "재벌"
 [9,] "반대"   "이명박" "경제"
[10,] "미국"   "안철수" "발표"
[11,] "이후"   "외교"   "개혁"
[12,] "결정"   "개헌"   "미국"
```

```
> #심상정
> terms(lda.out, 12)
      Topic 1  Topic 2    Topic 3
 [1,] "공약"   "안보"     "정부"
 [2,] "문제"   "국민"     "임금"
 [3,] "국민"   "문제"     "개혁"
 [4,] "복지"   "미국"     "국민"
 [5,] "전술"   "산업혁명" "청년"
 [6,] "정치"   "북한"     "국회"
 [7,] "해서"   "해서"     "일자리"
 [8,] "인상"   "산업"     "문제"
 [9,] "안철수" "병사"     "그것"
[10,] "민주"   "사람"     "노동"
[11,] "유승민" "북핵"     "민간"
[12,] "증세"   "상황"     "선거법"
```

```
> #홍준표
> terms(lda.out, 12)
      Topic 1  Topic 2    Topic 3
 [1,] "동성애" "거짓말"   "일자리"
 [2,] "사건"   "문재인"   "정부"
 [3,] "민주"   "정책"     "노조"
 [4,] "국민"   "유승민"   "노무현"
 [5,] "미국"   "박지원"   "강성"
 [6,] "검찰"   "지도자"   "귀족"
 [7,] "기업"   "검찰"     "청년"
 [8,] "수사"   "문제"     "기업"
 [9,] "사람"   "북한"     "투자"
[10,] "청와대" "중소기업" "홍준표"
[11,] "반대"   "햇볕"     "사퇴"
[12,] "안철수" "경찰"     "서민"
```

```
> #유승민
> terms(lda.out, 12)
      Topic 1  Topic 2    Topic 3
 [1,] "교육"   "북한"     "문제"
 [2,] "공약"   "문제"     "일자리"
 [3,] "정책"   "해결"     "정부"
 [4,] "국민"   "반대"     "북한"
 [5,] "대표"   "합의"     "문재인"
 [6,] "안철수" "개혁"     "공무원"
 [7,] "문재인" "핵미사일" "경제"
 [8,] "정부"   "전술"     "기권"
 [9,] "세금"   "중국"     "북한인"
[10,] "시간"   "재벌"     "사람"
[11,] "박근혜" "핵실험"   "안보"
[12,] "복지"   "한미"     "국민"
```

```
> #안철수
> terms(lda.out, 12)
      Topic 1 Topic 2    Topic 3
 [1,] "국민"  "교육"     "문제"
 [2,] "가지"  "가지"     "중소기업"
 [3,] "정부"  "해서"     "사람"
 [4,] "중국"  "문제"     "정부"
 [5,] "상황"  "국가"     "국민"
 [6,] "외교"  "산업혁명" "대북제재"
 [7,] "북한"  "정말"     "가지"
 [8,] "중요"  "시대"     "대기업"
 [9,] "안보"  "공약"     "상황"
[10,] "미래"  "가능"     "토론"
[11,] "세력"  "하나"     "변화"
[12,] "정말"  "정부"     "해결"
```

각 후보별로 LDA분석을 실시한 결과는 위와 같다. 먼저 문재인 후보자는 일자리, 지난 정권, 경제 성장 및 재벌 개혁과 같은 주제들을 주로 이야기했으나 중국, 외교, 핵실험과

관련한 외교 주제는 꾸준히 언급되었다. 또한 전체적인 단어들이 포괄적인 성격을 띈다. 심상정 후보자는 주로 복지 정책 공약, 북한과 안보, 노동 개혁 등과 관련한 이야기를 많이 했다. 홍준표 후보자는 동성애, 햇볕정책, 강성 노조 등의 다소 특색있는 단어들의 출현으로 토픽의 분류 기준이 흐릿해졌지만 전체적으로 검찰과 민주 국민, 각 후보자들과 정책 그리고 고용 경제에 관련한 이야기를 하였다. 유승민 후보자는 교육 등의 정책 공약, 북한, 북핵 등의 안보문제, 그리고 일자리 경제 등 토픽의 분류가 잘 되었다. 마지막으로 안철수 후보자는 외교, 교육 등의 공약, 그리고 기업, 국민, 대기업, 변화 등 기업 관련 정책들을 주로 이야기 하고 있음을 알 수 있다.

3.3.2. CTM

```
> #문재인
> terms(ctm.out, 12)
      Topic 1  Topic 2  Topic 3
 [1,] "안보"   "일자리" "정부"
 [2,] "국민"   "성장"   "합의"
 [3,] "박근혜" "경제"   "문제"
 [4,] "홍준표" "국민"   "반대"
 [5,] "나라"   "발표"   "안철수"
 [6,] "시간"   "재벌"   "미국"
 [7,] "정권"   "북한"   "중국"
 [8,] "해서"   "정책"   "북한"
 [9,] "개헌"   "재원"   "노무현"
[10,] "외교"   "해서"   "핵실험"
[11,] "북한"   "필요"   "퍼센트"
[12,] "심상"   "문제"   "이명박"
```

```
> #심상정
> terms(ctm.out, 12)
      Topic 1 Topic 2  Topic 3
 [1,] "안보"  "임금"   "공약"
 [2,] "문제"  "국민"   "개혁"
 [3,] "국민"  "정부"   "문제"
 [4,] "사람"  "비핵화" "복지"
 [5,] "정부"  "동맹"   "박근혜"
 [6,] "공약"  "미국"   "전술"
 [7,] "해서"  "북핵"   "증세"
 [8,] "지적"  "정치"   "정도"
 [9,] "시간"  "퍼센트" "국민"
[10,] "사회"  "해서"   "정부"
[11,] "청년"  "국익"   "해서"
[12,] "인상"  "경제"   "그것"
```

```
> #홍준표
> terms(ctm.out, 12)
      Topic 1    Topic 2  Topic 3
 [1,] "일자리"   "미국"   "청와대"
 [2,] "거짓말"   "국민"   "검찰"
 [3,] "정부"     "동성애" "노무현"
 [4,] "노조"     "정책"   "박지원"
 [5,] "문제"     "홍준표" "경찰"
 [6,] "청년"     "선거"   "문제"
 [7,] "강성"     "감사"   "인사"
 [8,] "귀족"     "하나"   "토론"
 [9,] "기업"     "민주"   "문재인"
[10,] "문재인"   "노무현" "동의"
[11,] "중소기업" "자유"   "해서"
[12,] "투자"     "달러"   "지난번"
```

```
> #유승민
> terms(ctm.out, 12)
      Topic 1    Topic 2  Topic 3
 [1,] "문제"     "교육"   "북한"
 [2,] "안보"     "일자리" "문제"
 [3,] "경제"     "공약"   "정부"
 [4,] "합의"     "정책"   "반대"
 [5,] "개혁"     "공무원" "대표"
 [6,] "시간"     "세금"   "국민"
 [7,] "국민"     "사람"   "문재인"
 [8,] "동맹"     "복지"   "노무현"
 [9,] "재벌"     "문제"   "안철수"
[10,] "기권"     "박근혜" "박지원"
[11,] "선제타격" "조원"   "핵미사일"
[12,] "중요"     "보수"   "전술"
```

```
> #안철수
> terms(ctm.out, 12)
      Topic 1 Topic 2    Topic 3
 [1,] "국민"  "문제"     "상황"
 [2,] "사람"  "중소기업" "교육"
 [3,] "가지"  "정부"     "가지"
 [4,] "세력"  "가지"     "정부"
 [5,] "북한"  "해서"     "산업혁명"
 [6,] "적폐"  "일자리"   "국가"
 [7,] "중국"  "대기업"   "변화"
 [8,] "중요"  "공약"     "대북제재"
 [9,] "토론"  "투자"     "시대"
[10,] "문제"  "위주"     "미래"
[11,] "정부"  "개혁"     "중요"
[12,] "정말"  "기술"     "안보"
```

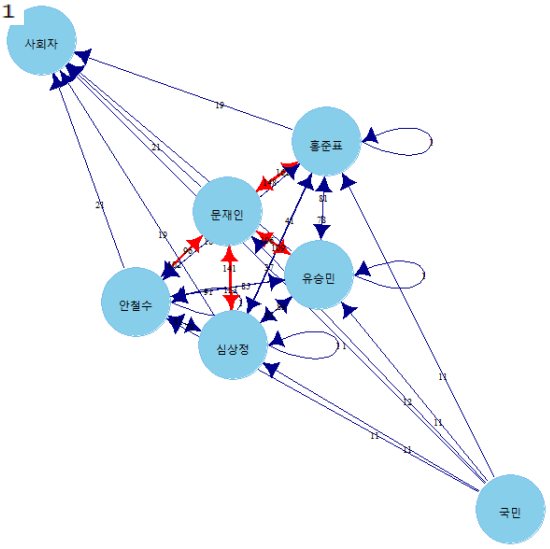
각 후보별로 CTM 분석을 한 결과는 위와 같다. 먼저 문재인 후보자는 안보와 보수정권 후보자, 일자리와 경제성장 정책, 그리고 대북 및 외교를, 심상정 후보자는 안보, 공약,사회,청년 등 전반적 정책 공약, 비핵화, 북핵 등 외교를, 홍준표 후보자는 일자리, 기업 정책, 정치인사를, 유승민 후보자는 안보, 경제 등 전반적 정책과 경제 및 고용 경제, 그리고 안보를, 안철수의원은 일자리, 기업, 개혁 등의 정책을 주로 이야기 했다고 나타났다.

따라서 문재인 후보자는 보수 정권, 경제 성장, 외교를, 심상정 후보자는 복지 정책 공약, 대북 외교와 안보, 홍준표 후보자는 기업 정책을, 유승민 후보자는 안보, 고용 경제를, 안철수 후보자는 기업 관련 정책과 관련하여 주로 토론했다고 할 수 있다.

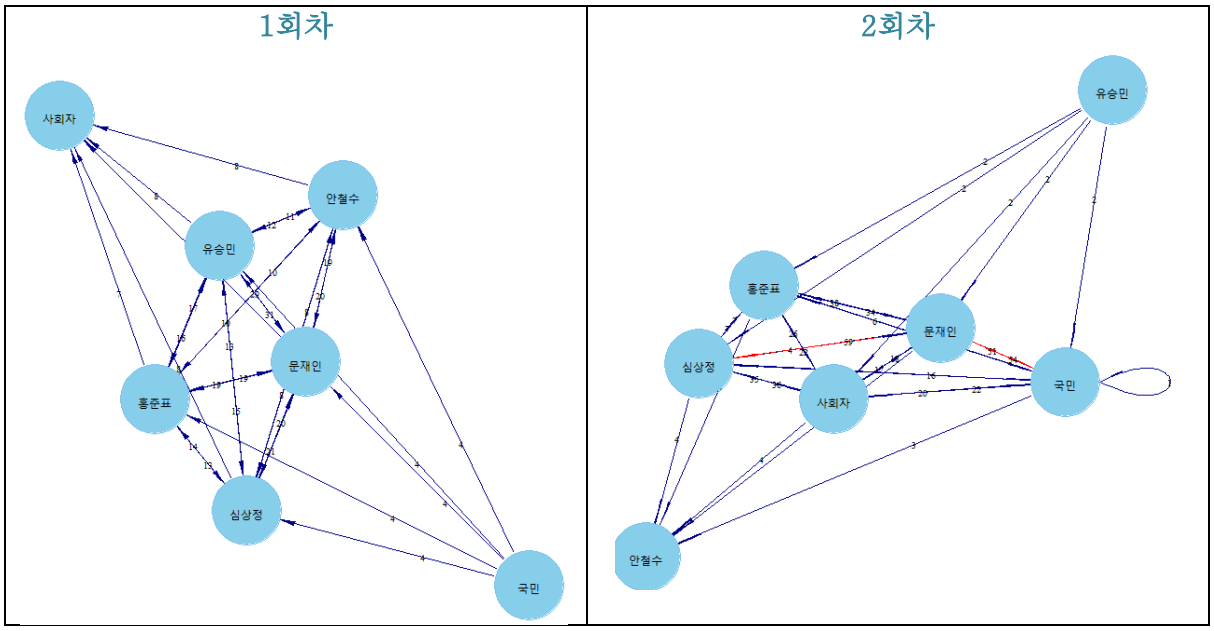
3.4 후보 별 토픽 네트워크 분석

```
> table(ft)
        To
From     국민 문재인 심상정 안철수 유승민 홍준표
  문재인   21      0     96    106    141    148
  사회자    1     12     11     11     11     11
  심상정   21    102      1     91     81     18
  안철수   19    109     85      1     89     78
  유승민   19    153     79     91      1     37
  홍준표   19    162     16     81     41      1
```

각 후보들에 대하여 누구에게 가장 많은 질문을 하고 받는지를 네트워크 그래프화 하였다. 우선, 왼쪽의 표를 보면 문재인 당시 대선 후보자의 발언 빈도가 타후보에 비해 높고 타 후보들로부터 많은 질문을 받았다. Igraph 패키지를 이용한 plot에서 문재인 후보자가 이 대선 토론의 중심이 되고 있음을 알 수 있었다. 여기서 빨간색으로 나타난 연결관계는 발언 빈도의 비율이 높은 노드만을 표시한 것이다. 여기서 각 노드 간 연결된 숫자는 발언 빈도를 의미한다.



각 회차에서의 토픽 네트워크 분석을 실시한 결과는 다음과 같다.



4. 추가 분석

4.1네거티브 토픽 분석

(1)네거티브 EDA – 각 회차별, 후보별 negative

	문재인	심상정	안철수	유승민	홍준표
1회차	1.01%	3.57%	1.27%	1.54%	7.58%
2회차	13.25%	7.5%	18.2%	6.4%	30.77%
3회차	6.41%	5.64%	10.14%	2.74%	10.26%
4회차	7.69%	7.47%	2.24%	5.13%	19.13%
전체	8.01%	6.37%	7.61%	4.47%	19.4%

위 표는 각 후보별 회차에 따른 전체 발언 중 negative 발언의 비율을 나타낸 것이다. 이 중 특히 홍준표 후보의 경우, 타 후보에 비해 negative 발언의 비율이 두드러지게 높다는 점을 확인할 수 있다. 2017년 한국갤럽조사연구소에 의해 실시된 실제 TV토론회 시청 후 홍준표 후보에 대한 인식 변화 조사에 따르면 TV토론회 시청 후 홍준표 후보에 대한 인식 변화는 '좋아졌다'가 18%, '나빠졌다'가 42%로 홍준표 후보의 이미지에 부정적인 영향을 미쳤다고 볼 수 있다. 반면 negative 발언의 비율이 가장 낮은 유승민 후보에 대한 조사는

‘좋아졌다’가 33%, ‘나빠졌다’가 16%로서 TV토론회가 유승민후보에게 긍정적 영향을 미쳤다고 볼 수 있다.

(2) LDA & CTM

Topic 1	Topic 2	Topic 3
"홍준표"	"미국"	"국민"
"사람"	"패망"	"토론"
"여성"	"희열"	"시간"
"당권"	"승리"	"적폐"
"스트롱맨"	"월남"	"나이"
"사과"	"패배"	"자유"
"세탁"	"전쟁"	"한국"
"박근혜"	"진실"	"석이"
"정지"	"논리"	"노무현"
"포스터"	"주도권"	"민주"
"선거"	"본인"	"양념"
"해서"	"공약"	"국장"

Topic 1	Topic 2	Topic 3
"미국"	"국민"	"시간"
"토론"	"당권"	"여성"
"세탁"	"노무현"	"사람"
"희열"	"적폐"	"사과"
"패망"	"유승민"	"스트롱맨"
"승리"	"박근혜"	"나이"
"패배"	"정지"	"누구"
"월남"	"선거"	"답변"
"전쟁"	"홍준표"	"설거지"
"기재"	"실패"	"얼마"
"민주"	"경남"	"공약"
"논리"	"자유"	"자리"

전체 토론내용 중 negative 발언의 토픽을 분석하기 위해 각각 LDA와 CTM을 이용한 결과이다. LDA의 첫 번째 토픽에서는 ‘홍준표’, ‘여성’, ‘스트롱맨’, ‘사과’ 등의 단어들이 나타났는데, 이는 홍준표 후보가 과거 발언했던 ‘스트롱맨’ 사건과 관련이 있다. 두 번째 토픽에서는 ‘미국’, ‘월남’, ‘패망’, ‘진실’, ‘승리’ 등을 통해 문재인 후보의 자서전과 관련된 내용을 엿볼 수 있다. 세 번째 토픽에서는 “양념”, “나이” 등의 단어들로부터 후보자들의 negative 발언과 관련이 있다고 할 수 있다.

CPM의 첫 번째 토픽에서는 LDA의 세 번째 토픽과 유사한 단어들의 집합을 볼 수 있으며, 따라서 문재인 후보의 자서전과 관련된 내용임을 추측할 수 있다. 두 번째 토픽에서는 ‘노무현’, ‘유승민’, ‘박근혜’, ‘홍준표’, ‘적폐’ 등을 보아 정당 및 후보자의 정치적 진영과 관련된 주제임을 알 수 있다. 세 번째 토픽은 LDA의 첫 번째 토픽과 같이 ‘스트롱맨’, ‘설거지’, ‘여성’, ‘사과’ 등 단어들을 통해 홍준표 후보의 스트롱맨 사건과 관련된 negative 발언이 주제로 나타남을 알 수 있다.

4.2 문재인 후보가 받은 질문 토픽 분석

2.1.2. To 변수의 EDA 결과와 같이 문재인 당시 후보자에게 가장 많은 질문이 쏟아졌다. 앞에서 살펴보았던 지지율과 결국 19대 대선 결과 문재인 후보자가 당선되었음을 바탕으로 보아, 가장 많은 질문을 받았다는 것은 가장 많은 견제를 받았다는 의미로 받아들일 수 있다. 따라서 본 연구에서는 문재인 후보자가 받았던 질문의 내용은 주로 무엇으로 구성되었는지에 대하여 분석해 보았다.

(1)EDA

```
> findFreqTerms(cpsTdm, lowfreq = 40)
[1] "공약"   "문제"   "정부"   "국민"   "북한"   "노무현" "토론"
[8] "일자리"
```

문재인 후보가 질문을 받는 만큼, 앞서 제거했던 단어들과 더불어 '문재인'이라는 단어 역시 무의미하다고 판단하여 제거하였다. 그 결과, 문재인 후보에게 했던 질문들에서 가장 많이 나타나는 단어들은 위와 같으며, 문재인 후보의 주 공약인 일자리 문제 해결 등에 대한 토론이 활발하였음을 알 수 있다.

```
> findAssocs(cpsTdm, '안보', 0.3)
$`안보`
      로운      위태      청산    디제이      수식
      0.58      0.58      0.58      0.58      0.58
      자금    퍼준거    동맹국      불안  안보문제
      0.58      0.58      0.58      0.58      0.58
      트랙    한국의      협력      목표    김정은
      0.58      0.58      0.58      0.51      0.41
      나라 바른정당하고    위기      위협      사정
      0.41      0.41      0.41      0.41      0.41
      중국      방향      진행      당선      도발
      0.39      0.33      0.33      0.33      0.33
    한반도      외교
      0.33      0.30
```

문재인 후보가 받은 질문들 중 '안보'와 함께 등장하는 비율이 30% 이상인 단어들에 대해 분석을 실시하였다. '불안', '안보문제', '위협' 등 국내외 부정적인 상황을 바탕으로 구성된 것을 확인할 수 있었다. 또한 '김정은', '중국' 등 대북 및 대중외교와 관련된 단어들도 상당수 등장하였다.

(2)LDA & CTM

```
     Topic 1  Topic 2      Topic 3
[1,] "일자리" "정부"       "북한"
[2,] "공약"   "공약"       "문제"
[3,] "국민"   "토론"       "정부"
[4,] "시간"   "정책"       "거짓말"
[5,] "사람"   "국민"       "핵실험"
[6,] "사건"   "문제"       "문재인"
[7,] "수사"   "노무현"     "노무현"
[8,] "청년"   "문재인"     "반대"
[9,] "필요"   "박근혜"     "중국"
[10,] "검찰"  "일자리"     "하면"
[11,] "공공"  "국가보안법" "국민"
[12,] "공무원" "증세"      "찬성"
```

```
     Topic 1  Topic 2  Topic 3
[1,] "정부"   "문제"   "일자리"
[2,] "문제"   "북한"   "공약"
[3,] "문재인" "핵실험" "국민"
[4,] "거짓말" "상황"   "정부"
[5,] "노무현" "국민"   "사람"
[6,] "북한"   "토론"   "복지"
[7,] "토론"   "민주"   "필요"
[8,] "정책"   "반대"   "사건"
[9,] "중요"   "미국"   "수사"
[10,] "박근혜" "개월"  "시간"
[11,] "재벌"  "중국"   "북한"
[12,] "대표"  "변화"   "책임"
```

문재인 후보가 받은 질문에 대해 LDA 및 CTM 분석을 실시하였다. LDA의 첫 번째 토픽에서는 '일자리', '사람', '청년' 등 정책 후보의 공약과 관련된 주제가 나타났다. 두 번째 토픽에서는 당시 정치적 의제에 관한 것이 주제로 나타났다. 세 번째 토픽에서는 '북한', '핵실험', '중국', '찬성', '반대' 등으로 보아 사드배치로 인한 외교 정책과 관련된 주제임을 알 수 있다.

CTM의 첫 번째 토픽에서는 문재인 후보에 대한 negative 발언과 연관된 단어들이 엿보인다. 두 번째 토픽에서는 '북한', '핵실험', '미국', '중국' 등으로 보아 사드 및 그와 관련한 외교적 상황에 대한 주제를 확인할 수 있다. 세 번째 토픽에서는 '일자리', '사람', '복지' 등 문 후보의 노동 및 사회복지 공약과 관련된 주제가 나타났다.

Ⅲ. 결론

19대 대선 토론회의 내용을 바탕으로 텍스트 분석을 실시한 결과, 후보별, 분야별로 나누어 토론회를 면밀하게 분석할 수 있었다. LDA 및 CTM 기법을 활용하여 토론회의 주된 의제를 한 눈에 살펴볼 수 있었다. 본 연구에서 드러난 각 후보 별 주요 토론 내용 및 토론회 직후의 지지율 변화를 비교하여 어떤 의제가 후보자에게 불리하게 또는 유리하게 작용하였는지를 간략하게나마 살펴볼 수 있었다.

서론에서 제시되었던 대선토론회 다음날의 지지율 변화 추이를 바탕으로 회차별 토론 내용 또는 진행 방식에 따라서 지지율에 영향을 미쳤다고 판단할 수 있었다. 특히, 가장 지지율을 높았던 후보를 제외하고, 네거티브 발언에 지지율이 영향을 많이 받았다는 것을 알 수 있었다. 네거티브는 토론의 주된 의제가 아님에도 불구하고, 토픽 분석을 실시하였을 때, 한 토픽이 네거티브 관련 내용의 토픽으로 분류될 수 있을 만큼 매 토론에서 상당한 비중을 차지한 것을 알 수 있었다.

그러나 이러한 분석을 시행하는 것에도 한계점은 존재했다. 후보자들의 지지율의 변동에는 토론회외에도 기타 상황적 변수들이 관여했지만, 주어진 데이터만으로는 외부적인 파생변수를 통제하기 어려웠기 때문이다. 또한, 텍스트 마이닝 기법을 이용할 때 나타나는 방법론적 한계가 이 분석에서도 발생했다. 텍스트라는 질적 데이터를 정형화된 분석틀에서 연구함으로써 질적 데이터가 가지는 고유한 의미를 함께 활용해 추론하는 것에 한계점이 나타났기 때문이다. 또한, 토론의 어떠한 요소가 지지율 차이에 영향을 미쳤는지 정확하게 특정하는 데에도 한계를 보였다.