국가전략, 제26권 4호 2020년 겨울호, 85~119.

빅데이터를 활용한 박근혜·문재인 정부 간 대외정책 토픽 비교 분석: 외교부 장관 연설문을 중심으로

이 시 온*

❖ 요 약 ❖

본 연구의 목적은 박근혜 정부에서 문재인 정부로의 정권교체에 따른 대외정책 변화를 정부별 외교부 장관 연설문 텍스트마이닝을 통해 객관적이고, 정량적으로 비교 분석하는 것이다. 연구의 주요 결과는 다음과 같다. 우선 정부 간 공통 빈출 키워드는 북한, 한반도, 평화 등이며, 박근혜 정부의 특징적 단어는 동북아, 경제, 안보, 통일 등인 반면, 문재인 정부는 국민, 인권, 지속, 대화 등의 단어를 주로 사용하였다. 다음으로 토픽모델링 분석을 통해 박근혜 정부의 8개(① 한반도 평화통일과 유라시아 이니셔티브, ② 국민행복시대를 열어가는 외교, ③ 국익을 위한 외교·안보 전략, ④ 북한 인권문제 및 안보 위협에 따른 국제사회 대응, ⑤ 경제 협력 및 교류 확대로 파트너십 확대, ⑥ 국제사회 내 한국의 주요 현안 해결 노력, ⑦ SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할, ⑧ 동북아 평화협력 구상) 주요 토픽을 발견하였고, 문재인 정부에서는 6개(① 평창올림픽 계기 대화를 통한 북핵 문제 해결, ② 국가 및 지역 간 경제 협력 증진을 통한 발전 모색, ③ SDGs 달성을 위한 국제사회 지원 노력, ④ 한반도 및 세계평화를 위한 남북 및 북미 간 정상회담, ⑤ 유엔 내 한국의 여성·평화·인권 문제 기여 노력, ⑥ 국민과 소통하는 외교)의 대표적 대외정책 과제를 도출하였다. 또한, 시간의 흐름에 따라 외교 이슈 간 비중의 변화를 파악함으로써 특정 정부 내에서도 시기별로 강조하는 외교정책의 우선순위가 달라짐을 실증적으로 확인하였다.

핵심어: 대외정책, 외교전략, 텍스트마이닝, 빅데이터분석, 토픽모델링, 잠재디리클레 할당, LDA

DOI: 10.35390/sejong.26.4.202011.004
* 서울대학교 행정대학원 박사과정

Ⅰ. 서론

본 연구에서는 헌정 사상 최초의 대통령 탄핵 및 이후 여야 간 정권교체라는 특수한 배경하에서 등장한 박근혜 정부와 문재인 정부 간 대외정책의 차이를 실증적으로 비교 분석하고자 한다. 기존에도 국내 정권 변화에 따른 대외정책의 차이를 비교 분석하는 선행연구는 많이 찾아볼 수 있었으나, 대개 특정한 사례 중심으로 연구를 수행하고 있어 정부별 대외정책을 전반적으로 검토하고 비교하는 연구는 상대적으로 부족하였다. 동시에 외교정책연구의 전반적인 기조가 국제관계이론에 근간을 두고 있으며, 과학적, 통계적 접근을 통한 데이터 기반 연구 축적은 상대적으로 미약한 것으로 평가되었다(김일수 2015). 일부 실증적 자료에 기반한 연구가 있다 해도 대다수가 소수 외교관의 경험적 지식이나 개인적 네트워크를 활용한 스몰데이터 분석에 그치고 있다는 지적(김상배 2017) 역시 있었다. 이러한 인식에서 동 연구는 빅데이터를 기반으로 한 국내 외교정책 분석의 가능성을 외교부 장관의 연설문 텍스트마이닝(Text mining)을 통해 살펴보고, 정부별 핵심 대외정책 차이에 대한 이해도를 높임과 동시에 정부 내 세부 정책 간 상대적인 우선순위와 중요도를 파악하여 향후 정부의 대외정책 수립에 시사점을 제공하고자 한다.

텍스트마이닝이란 비정형 텍스트 데이터에 대한 자연어처리(Natural language processing: NLP) 기술을 기반으로 단어 사이 숨겨진 관계나 패턴을 정량화하여, 의미 있고 활용 가치 높은 정보로 가공하는 기법이다(Preiss and Stevenson et al 2015). 근 10년 사이 마케팅 전략, 고객관리 등의 민간 차원뿐 아니라 학술연구 및 공공정책 개발에서도 빅데이터 분석은 점차 그 활용도가 확대되는 추세이나, 대외정책연구 분야에서는 상대적으로 아직 낯선 영역으로 남아있다. 이번 연구는 다양한 텍스트마이닝 기법 중에서도 LDA(Latent Dirichlet Allocation, 잠재디리클레 할당) 방식을 활용한 토픽모델링(Topic modeling)을 실시함으로써 연설문에 드러난 한국 외교정책의 변화 추이를 계량화하는 것을 목표로 한다. 특정 국가의 대외정책 검토를 위해서는 보도자료, 발간 도서, 기고문, 인터뷰, 녹취록 등 다양한 기록물을 고려할 수 있으나, 외교정책에서 외교부 장관이 가지는 역할 및 직위[1]를

1) (정부조직법 제30조 제1항) 외교부 장관은 외교, 경제외교 및 국제경제협력외교, 국제관계 업무에 관한 조정, 조약 기타 국제협정, 재외국민의 보호·지원, 재외동포정책의 수립, 국제정세의 조사·

고려할 때, 장관 연설문은 한 국가의 외교정책이 지향하는 가치와 외교 우선순위를 확인할 수 있는 대표적인 공식 정보 원천이라 판단하여 본고는 외교부 장관 연설문을 분석에 활용하였다. 외교부 장관의 임명권자는 국가 최고 의사결정권자인 대통령이며, 정권 변화는 곧 장관 교체를 동반한다는 점에서 장관은 대통령의 대리인이자 외교부의 총 책임자로 해당 정부의 외교 방향을 결정하고, 공식적으로 발표할 권한 및 책임을 동시에 지닌다. 따라서 장관 연설문을 특정 개인의 견해나 인식의 표현만으로 평가하기보다 특정 정부를 대표하는 대외정책 공식 발표 창구로 간주하는 것이 적절하다고 판단하였다.

결론적으로 본 연구는 정부 교체에 따른 정부 간 대외정책 키워드 및 정책의제에 차이가 있음을 밝혀냈다. 먼저 두 정부 모두 북한, 한반도, 평화 등의 키워드를 주로 사용하고 있었으나, 여타 핵심 어휘에서 차이를 보였는데 우선 박근혜 정부에서 특징적으로 나타난 단어는 동북아, 경제, 안보, 통일 등이고, 문재인 정부에서는 국민, 인권, 지속, 대화 등의 단어 빈도가 높게 나타나 정부 간 대외정책 관심사에 차이가 있음을 알 수 있었다. 또한, 토픽모델링 분석을 통해 박근혜 및 문재인 정부의 대외정책의제를 분석한 결과, 박근혜 정부에서는 북한, 한반도, 동북아, 국익, 국제사회, 경제, 국민 등 7개 주제, 문재인 정부는 북한, 한반도, 국제사회, 경제, 국민 등 5개 주제를 중점적으로 다루고 있었다. 일부 공통된 정책 대상이 있었으나, 세부 토픽에는 상당한 차이를 보여 정부에 따른 대외정책 접근 방식의 변화를 알 수 있었고, 더불어 특정 정부 내에서도 시기별로 토픽에 대한 우선순위에 차이가 발생함을 확인함으로써 정부 간 중점 대외정책의제를 효과적으로 비교 분석하였다.

본 연구의 구성은 다음과 같다. 서론에 이은 2장에서는 기존 선행연구를 비판적으로 검토하고, 그 속에서 빅데이터 기반 한국 외교정책 연구가 갖는 의미를 살펴보고자 한다. 3장에서는 외교백서를 기반으로 텍스트마이닝 분석을 위한 핵심적인 외교 주제를 설정한다. 이를 기반으로 4·5장에서는 빈도분석(Term Frequency Analysis), 토픽모델링(Topic Modeling) 등의 텍스트마이닝 기법을 활용하여 연구를 설계하고, 분석을 수행한다. 6장에서는 결론적으로 분석 내용을 요약정리하고, 결과를 토대로 정책적 시사점을 제시하고자 한다.

분석에 관한 사무를 관장한다.

Ⅱ. 선행연구 검토 및 이론적 논의

1. 외교정책 결정요인 및 연설문 분석의 함의

외교정책 혹은 대외정책(foreign policy)이란 한 국가가 국익을 증진하고, 보호하기 위하여 상대국가 또는 국제사회에 대해 행하는 국가 차원의 행동 전략을 의미한다(이수진·이신화 2019, 조명현 2000). 관련하여 외교정책 요인 연구의 대표적인 학자 몇 사람을 살펴보자면 다음과 같다. 예를 들어 월츠(Waltz 1959)는 외교정책의 가장 극단적인 형태인 전쟁을 기반으로 원인을 개인, 국가, 국제체제 차원에서 분석할 수 있다고 주장하였고, 로즈나우(Rosenau 1971)는 외교정책 결정요인을 개인, 역할, 정부, 사회, 체제의 5개 수준으로 유형화할 수 있으며, 각 요인의 상대적인 중요성은 한 국가의 영토 크기와 자원 보유량, 경제발전수준, 체제 개방성 등에 따라 달라진다고 설명하였다(Rosenau 1976). 한편, 앨리슨(Allison 1999)은 쿠바 미사일 위기 사례 분석을 통해 외교정책 결정을 '합리적 행위자', '조직절차', '관료정치' 등 세 가지 모델로 나누어 제시하고 있다.

그 밖에도 외교정책에 영향을 미치는 요소에 관한 연구는 비교적 활발히 이루어져 왔음에도 상대적으로 체계화된 이론이 발전되지 못한 것은 근본적으로 외교정책이 가진 다각적이고 포괄적인 측면 때문이다. 따라서 어떤 측면만이 중요하다고 강조하는 것은 상대적으로 다른 부분이 간과되고 있다는 의미로 해석될 수도 있다.

따라서 동 연구는 양 정부의 정책 차이의 원인을 단순히 내부적으로 대통령 1인의 리더십 변화 또는 세계질서 등 외부적 환경 변화에 따른 반응 등으로 나누어 구분하여 해석하는 대신, 정책을 유사한 정치적 입장과 인식을 공유하는 대통령-장관(행정부)-국회(집권여당)의 공동 산출물로 간주하고 분석하고자 한다. 같은 맥락에서 분석의 대상이 되는 외교부 장관의 연설문 역시 한 개인의 특성이나 리더십 스타일적인 면에서 바라보기보다, 한 정부의 충실한 대리인이자 대표에 의한 공식적인 대외정책 방향을 표명한 것으로 보고 해석할 수 있다.

연설문은 정책 분석을 위한 효과적인 도구 중 하나로 알려진 만큼 기존에도 연설문을 기반으로 정책의 우선순위 및 의제설정 방향을 분석한 다수의 선행연구가 있다. 이창길(2010)은 노무현 및 이명박 대통령의 연설문 간 내용분석을 통해 정부별 가치

지향 및 정책의제설정의 차이를 살펴보았고, 대통령 연설과 정책승인 간 강한 상관관계를 확인한 연구(Strömbäck and Kiousis 2010) 및 1989년~2000년 사이 미국 대통령 연설문이 대통령 제안 정책의 입법화에 미치는 기여(Eshbaugh-Soha 2010)를 살펴본 연구를 통해서는 연설문과 정책의 밀접한 관련성을 확인할 수 있었다. 연설문을 기반으로 대통령의 정책의제설정 방식(김혁 2016)을 연구한 결과 역시 연설문이 정책의제설정에 상당 부분 기여함을 보여주었다.

대통령 이외에도 주요 정책 의사결정자의 연설문에 기반한 정책 분석 연구는 다수 있으며 대표적인 사례를 소개하자면 다음과 같다. 우선 김미숙과 전상완(2019)은 문화체육관광부 장관 취임사를 바탕으로 역대 장관별 스포츠 정책 변화를 살펴보고, 향후 스포츠 정책에 대한 시사점을 도출하였다. 한편, 미국 통화·금리정책 결정기구인 연방공개시장위원회(FOMC: Federal Open Market Committee) 소속 위원의 연설문을 바탕으로 통화정책을 분석한 전영환(2014)의 연구도 찾아볼 수 있다.

한편, 김상배(2017)는 특별히 외교안보 분야 빅데이터 분석을 할 때 유력인사 기반 접근(influencer-based approach) 즉, 일반 대중이 아닌, 외교 이슈에 영향을 미칠 수 있는 인물 및 기관의 담론 분석을 통한 접근법이 필요하다고 지적하고 있다. 이상의 선행연구 결과에 의하면 정부 간 외교정책 분석을 위해 외교부 장관의 연설문을 활용하는 것은 타당한 접근 방식으로 판단된다.

물론 모든 정책의 최종 결정권자인 대통령의 권한을 강조하는 입장에서는 장관의 담화문이 가지는 의미를 축소할 수도 있다. 그러나 박근혜 정부와 문재인 정부 시기 정책의 차이를 단순히 대통령의 개인적 성향 차이에 따른 정책 노선 변경이라고만 보기 어려운 것은 같은 시기 집권 여당이 국회 의석수를 절반 이상 차지하였고, 행정부 전반의 대대적인 인력교체 역시 동시에 발생했다는 점에서 정책 변화의 요인은 다각도로 파악되어야 한다. 한편, 정무직으로서 평균 1년 내외의 짧은 임기를 가지는 여타 부처의 장관 대비, 윤병세, 강경화 장관은 각 정부 출범 이후로 단 한 번의 교체도 없이 장관이 외교정책을 담당했다는 점(단, 강경화 장관의 경우 동 연구 진행 시점 기준)에서 공통으로 두 장관에 대해 대통령과 집권 여당, 외교부 내부 등 안팎의 신뢰와 지지가 뒷받침되었다는 해석에 무게가 실린다. 또한, 연구 주제가 대외정책에 국한된다는 점에서 대통령 연설문 대비, 해당 분야만을 다루는

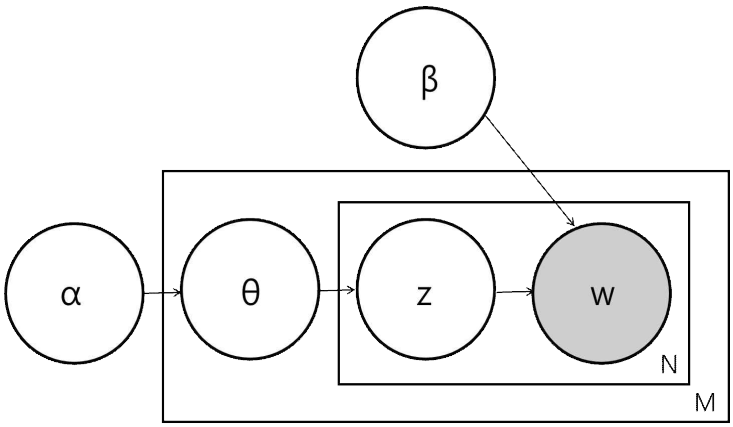
장관 연설문을 분석하는 것이 분석대상의 양과 내용적인 깊이에서 장점을 지닌다고 판단하였다.

2. 토픽모델링(Topic Modeling)을 활용한 정책연구

토픽모델링이란 대량 문서에 내재한 단어를 근간으로 토픽을 추출하고, 토픽 간 관계 분석을 통한 변화를 파악하는데 유용한 빅데이터 기반 텍스트분석 기법이다(김지은 2017). 또한, 토픽 분류 결과를 살펴봄으로써 문서가 지닌 특징적인 주제에 관한 정보를 구할 수 있다. 토픽모델링 중에서도 최근에 활발히 사용되는 잠재디리클레할당(LDA) 알고리즘은 각 문서 내 존재하는 복수(k)의 주제가 나타날 확률을 기반으로 특정 문서의 특성을 나타낸다(Blei and Ng et al 2003). 여기서 k는 전체 문서 집합을 가장 잘 표현할 수 있는 숨은 변인(hidden variable)의 수로서, 반복적으로 모델을 생성한 후, 통계적으로 비교하여 적정값을 찾거나, 혹은 여러 k 값의 결과에 대한 연구자의 직관적 판단으로 값을 찾을 수도 있다(김영욱·함승경 외 2017). 두 가지 방법 모두 장단점이 있으나 본 연구에서는 연구자가 외교백서를 근거로 전체 연설문에 대한 사전 검토 후, 반복적인 테스트를 통해 최적 토픽 수를 설정함으로써 가능한 객관적인 판단 기준을 견지하려고 노력하였다.

다음 그림은 LDA에 대한 개념도로서 수집된 M개의 문서(N) 안에는 단어(w)가 존재하며, 각 단어에는 대응되는 토픽(z)가 있다고 가정할 때, 실제 각 문서에는 여러 토픽이 내재되어 섞여 있고, 문서에 따라 토픽분포가 다르다고 할 수 있다. 여기서, α는 디리클레 분포의 비율 파라미터이고, θ_i는 각 문서의 토픽 비율을 의미한다.

〈그림 1〉 LDA(Latent Dirichlet Allocation) 개념도



출처: Blei and Ng et al(2003)를 참조하여 필자 작성

즉, 하나의 문서는 다양한 주제(토픽)의 조합으로 구성되며, 각 문서의 토픽은 다변량 주제 분포에 의해서 결정된다. 예를 들어 문서가 10개의 주제로 구성되어 있다고 하면, 주제 다변량 θ_i는 M개의 요소를 가지고 있는 θ_i=[θ_{i1},θ_{i2},...,θ_{iM}] 확률변수이고, θ_{i1}+θ_{i2}+...+θ_{iM} =1을 만족한다. 만일, 한 문서에 100개의 단어가 있다면 100개의 각 단어의 위치에 각 주제를 θ_i로부터 생성하며, I번째 문서의 j번째 주제는 Z_{ij}라 한다. 이 주제 Z_{ij}에 의하여 단어가 결정되는데, 이는 각 주제에 대한 단어 분포에 의한다. 즉, 주제 k는 단어를 산출하는 각각의 확률을 가지게 되는데, 이를 Φ_k라고 하고, 이때 단어의 수가 V개라면 Φ_k=[Φ_{k1},Φ_{k2},...,Φ_{kV}]가 된다. 다변량 확률변수 역시 디리클레 분포를 따르며, Φ_{k1}+Φ_{k2}+...+Φ_{kV}=1을 만족하게 되는데, 결론적으로 Φ_k에 의하여 최종적인 단어 w_{ij}가 결정되는 것이다.

지금까지의 빅데이터 활용 정책연구는 대체로 텍스트마이닝 기법에 근거한 경우가 많고, 다양한 세부 정책 분야에서 등장한다. 먼저 송태민(2013)은 자살 방지 정책 개발을 위해 SNS 상 자살과 관련된 검색어를 수집하여 빅데이터 분석을 실시하였고, 한국정보화진흥원(2020) 역시 청소년 자살 예방정책 수립을 위해 인터넷 뉴스, 블로그, 카페, SNS 등 다양한 채널에서 자살 관련 키워드를 분석하였다. 한편, 김정미와 윤미영(2012)은 기존 보육정책에 대한 평가를 위해 SNS 기반 감성분석

및 키워드 상관분석, 내용분석 등의 분석을 수행한 반면, 권기석과 박지수 등(2014)은 네이버 크롤링으로 자료를 수집하고 교육정책 연구 논문과의 비교를 통해 실제 교육정책 수요를 파악하였다. 한편, 임화진과 박성현(2015)은 지방자치단체의 지역 정책 수립을 위해 언론기사를 수집하고, 단순빈도 및 TF-IDF 빈도분석을 통해 정책 의제설정의 시사점을 발굴하였고, 끝으로 정예림과 강정은(2019)은 SNS 분석을 기초로 기후변화와 연관된 대중들의 정책 수요를 발굴하고자 하였다. 이상 연구를 살펴보면 텍스트마이닝의 주요 정보 원천은 인터넷 뉴스, 신문 및 SNS 등이 주로 활용되며, 연구 주제는 정책 수요 및 동향 분석을 중심으로 발전하고 있음을 알 수 있다.

텍스트마이닝 중에서도 빈도분석 등 단어 단위 분석의 한계를 극복하고, 전체 문서 내 주제를 분석하는 토픽모델링을 활용한 정책 관련 연구도 일부 찾아볼 수 있는데, 몇 가지 사례만 소개하면 다음과 같다. 우선 홍영희(2019)는 소프트웨어 교육정책을 정책형성, 결정, 집행 단계 등으로 나누어 단계별 주요 토픽을 도출하였고, 김지은(2017)은 대학구조개혁 평가에 대한 교육부 문서 및 신문 기사의 토픽을 추출하여 양 토픽 간 공통점과 차이점을 분석하였다. 김형지와 김선우 등(2018)은 스마트폰 중독에 대처하기 위한 정부 정책 비교를 위해 2010~2018년 사이 보도자료를 토픽화하고, 정책 시기별로 주요 쟁점을 파악하였다.

그러나 다른 정책 연구 분야와 달리 상대적으로 외교정책 분야에서 토픽모델링 활용 연구는 아직 흔하게 활용되는 분석 기법은 아니다. 대표적으로 사드(THAAD) 이슈에 대한 한·중 언론의 태도를 비교하기 위해 양국 신문 기사 토픽모델링을 실시한 정원준(2018)의 연구가 있었고, 한국과 미국의 정권교체 시기 외교전문의 주제 변화를 분석한 연구(백영민·최문호 외 2014) 정도가 외교 이슈와 관련된 토픽모델링 연구로서 이를 통해 외교 분야도 충분히 텍스트마이닝 및 토픽모델링 분석의 대상이 될 수 있다는 가능성을 엿볼 수 있다.

Ⅲ. 외교백서를 통해 살펴본 한국 대외정책의 주요 과제

1. 외교백서의 의의

외교백서((外交白書)란 1990년 이후 매년 한국 정부가 발간하는 연간 국제정세, 정부 외교정책 기조와 주요 외교활동을 종합 수록한 책자이다(외교부 2020). 백서(white paper)의 기원은 영국 내각으로 정부가 특정 정책에 대한 공식 입장이나 조사 결과를 정리하여 의회에 제출하는 흰색 표지의 보고서를 의미한다(서호준 2019). 본고에서는 외교백서에 드러난 공식적인 한국의 외교정책 및 추진 성과를 살펴봄으로써 기간별 대외정책 흐름을 알고, 대외정책의 주요 과제를 파악하여 정부 간 정책 비교 항목을 설정하는 기초로 활용하고자 한다.

2. 외교백서를 통해 살펴본 정부 간 대외정책 비교

박근혜 정부 및 문재인 정부 시기인 2013년부터 2019년까지의 외교백서는 총 7권이며, 각각은 총 7장 및 부록으로 구성되어 있다. 특히 제1장에는 해당 연도의 국제정세 및 외교정책 기조가 제시되어 있어 전반적인 정책의제 파악에 도움을 준다. 이에 따르면 전 기간 외교백서 7권에 공통으로 등장하는 한국의 대외정책 대상은 총 4가지로 요약 가능하다.

첫 번째, 한국 대외정책에서 공통적인 최우선 과제는 북한이다. 두 번째, 양자관계에서는 한반도를 둘러싼 주요 열강, 즉 미국, 중국, 일본, 러시아 등 4개국과의 외교가 매우 중요하게 다루어지고 있었다. 다음으로 지역별 또는 다자관계를 고려한 외교정책이 수립되어 추진되고 있고, 끝으로 한국민을 대상으로 하는 전략이 주요 대외정책 내 포함된다.

다만 정부가 표방하는 외교 관련 국정 목표에는 차이를 보였다. 박근혜 정부의 핵심 외교 가치는 '신뢰'이고, 추진 방식으로 '국민과 함께하는 외교'를 전면에 내세우고 있었다(외교부 2014, 2015, 2016). 반면, 문재인 정부는 '평화'와 '번영'이라는 두 핵심가치를 축으로 대외정책의 목표 지역을 '한반도'로 특정하고 있다는 차별점을 나타냈다.

양 정부는 세부적인 외교 과제에서도 차이점을 보였는데 북한 관계에서 박근해 정부의 중점과제는 '북핵 문제 진전을 위한 동력 강화'이다. 즉, 박근혜 정부는 대북 문제의 핵심을 북핵으로 특정하고 있었다. 또한, '평화통일'을 전면에 내세움으로써 북한 이슈의 궁극적인 지향점을 통일로 설정하고 있음을 명시하였다. 이와 달리 문재인 정부는 북한을 직접적인 정책 대상으로 언급하기보다 '한반도 평화 프로세스 진전'이라는 공동 목표 달성을 위한 파트너로 보고 접근하고 있었다.

이어서 박근혜 정부의 주변국 외교 목표를 살펴보면 미국에 대해서는 한미동맹을, 중국에 대해서는 한중 동반자 관계의 조화 및 발전을, 일본에 대해서는 한일관계 안정화를 구체적으로 내세우고 있다. 또한, 동북아 평화협력 구상과 유라시아 협력 확대 등을 주요 외교정책 목표로 설정하고 있으며, 한편, 문재인 정부에서는 '전략적 로드맵에 따른 주변국 외교'라는 모토 하에 주변 국가와의 관계 개선을 도모하고자 하였다. 특히, 주변국에 대해서는 '협력외교'라는 명칭에서 알 수 있듯 미국, 일본, 중국, 러시아와의 관계 개선 및 강화에 초점을 맞추고 있었다. 또한, '외교 다변화를 통한 외교 지평 확대'라는 정책 목표 아래 주변 4강 이외의 국가 및 지역과의 외교가 중요하다는 점을 강조함으로써 신남방·신북방정책 등 대외관계의 확장을 꾀하였다.

다음으로 다자협력 혹은 글로벌 이슈에 대해서는 상대적으로 양측의 차이를 크게 발견하기 어려웠다. 박근혜 정부에서는 다자외교의 목표를 '세계평화와 발전에 기여하는 책임 있는 중견국'으로 설정함으로써 한국의 대외적 위상에 걸맞은 국제적 기여에 초점을 맞추었다면 문재인 정부에서는 '국제사회 내 역할 확대와 위상 제고'로 문구상 큰 차이를 보이지 않았다.

다음 대국민 목표로 박근혜 정부에서는 '재외국민 안전·권익 보호와 공공외교 확대'를 설정함으로써 전통적 외교에서 공공외교[2]로 외교의 개념이 확장되고 있음을 확인하였다. 국민 중에서도 특히 재외국민을 정책 대상으로 강조하고 있었는데, 문재인 정부에서도 유사하게 '국민과 함께하는 외교 실현'을 대국민 외교 과제로 설정하였으나, 구체적으로 재외국민보다 한 발짝 더 나아가 전체 국민을 외교의 주체로 보고 있음을 나타내었다.

---

2) 공공외교(public diplomacy)는 전통적 외교와 구분되는 국제관계의 새로운 패러다임으로 소프트파워(soft power)를 통한 새로운 관계의 힘을 사용하여 타국과 관계를 맺고 의사소통하면서 궁극적으로 자국의 이익을 꾀하는 외교 행위이다(백우열 2017).

끝으로 박근혜 정부의 경우, 기존 외교통상부가 외교부로 변화함에 따라 통상영역이 업무 범위에서 제외되었음에도 '경제외교를 통한 경제부흥 선도'라는 정책 목표를 명시한 것은 외교정책에서 대외경제정책이 차지하는 비중을 시사한다.

아래의 표는 이상의 정부별 대외정책 과제 및 세부목표를 정리한 것이다.

〈표 1〉 2014~2019년 외교백서에 나타난 정부별 대외정책 과제 및 세부목표

정부	국정목표	정책 과제					
		① 북한	② 주변국	③ 지역	④ 국제사회	⑤ 국민	⑥ 경제
박근혜	국민과 함께하는 신뢰 외교	북핵문제 진전 및 평화통일	전략적 로드맵에 따른 주변국 외교	동북아 평화협력 구상과 유라시아 협력 확대	세계평화와 발전에 기여하는 책임 있는 중견국 실현	재외국민 안전·권익 보호와 공공외교 확대	경제외교를 통한 경제부흥 선도
문재인	평화와 번영의 한반도	한반도 평화프로세스 진전	주변 4국과 협력외교 강화	외교 다변화를 통한 외교 지평 확대	국제사회 내 역할 확대와 위상 제고	국민과 함께하는 외교 실현	-

다음 장에서는 상기 표에서 정리된 내용을 바탕으로 외교백서에 나타난 정부의 국정목표 및 정책과제가 외교부 장관의 연설문 텍스트마이닝 분석결과에도 동일하게 나타나는지 살펴보고자 한다. 또한, 외교백서 상 드러난 정부간 대외정책 기조의 차이가 연설문에도 반영되고 있는지 혹은 연설문을 통해 추가적인 차이를 발견할 수 있는지를 분석함으로 박근혜 정부와 문재인 정부의 외교정책을 비교·대조하고 정책적 시사점을 찾고자 한다.

Ⅳ. 연구설계

1. 연구대상 및 자료수집

본 연구의 분석대상은 박근혜 정부의 윤병세 장관 연설문 213건 및 문재인 정부의 강경화 장관 연설문 42건 등 약 7년간(2013년 3월 13일에서 2020년 5월 31일) 발표된 외교부 공식 국문연설문[3] 총 255건이다. 다만, 연구 진행 시점 기준 문재인 정부의 임기가 2년가량 남아있는 점을 고려할 때, 정부 간 분석자료의 정보량에 차이가 있다. 또한, 2020년 2월 이후에는 연설문이 업로드 되고 있지 않은데, 이는 동 시기 세계적으로 확산된 코로나19의 영향으로 국가 간 이동이 대폭 제한됨에 따라, 장관 연설의 기회 자체가 없었을 것으로 추측된다. 이에 따라 동 연구 분석대상인 연설문은 2020년 2월 코로나19 확산 이전으로 제한되었고, 코로나19 이후 새롭게 등장한 정책과제인 코로나 대응을 위한 전 세계 공조 또는 K-방역 등의 외교 이슈는 아쉽게도 동 분석대상에 포함되지 않았다.

〈표 2〉 연구 분석대상

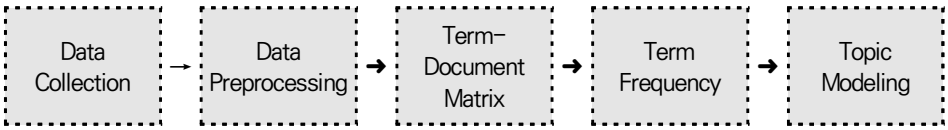
부처명	정부명	역대	장관명	임기	분석대상
외교부	박근혜 정부	37대	윤병세	2013년 03월 13일 ~ 2017년 06월 18일	213건
	문재인 정부	38대	강경화	2017년 06월 18일 ~ 현재	42건
연설문 총계					255건

3) 연구의 기초 데이터가 되는 장관 연설문 자료는 오픈소스 데이터 분석 도구인 R의 웹 스크래핑(Web scraping) 기법을 활용하여 외교부 공식 웹사이트(www.mofa.go.kr/minister/brd)에서 수집 및 취합하였다. 한편, 통상적으로 영문 연설문의 경우에 외교부 차원의 국문 번역본이 제공됨을 고려하여, 일관성 있는 분석을 위해 국문으로 한정하여 분석을 시행하였다. 따라서 예외적으로 번역본을 제공하지 않은 윤병세 장관의 연설문 17건과 강경화 장관의 연설문 5건 등 총 22건의 연설문은 금번 연구대상에서 제외되었다.

2. 분석 방법

동 연구는 아래와 같은 절차로 수행되었다. 수집된 다량의 텍스트(연설문)는 불용어(stop words) 처리, 형태소 분석 등의 전처리(preprocessing) 과정을 거친다. 그 후 단어-문서 행렬을 구성하여 빈도분석을 실시하고, 토픽모델링 기법을 통해 전체 연설문을 관통하는 복수의 주제 및 주제별 주요 단어를 파악한다. 끝으로 토픽 시계열 분석을 통해 시간의 흐름에 따라 대외정책 현안의 변화를 확인함으로써 연구를 종료한다.

〈표 3〉 연구절차



1) 전처리(preprocessing)

본 연구에서는 전처리 과정으로 수집된 총 255편의 연설문 중 한글 이외의 언어 및 특수문자, 숫자 등을 제거하고, 단어 정규화(word normalization) 과정을 거쳤다. 이후, 두 글자 이상의 단어에 대해서만 의미를 내포한 주제어로 판단하고 추출하였다. 다만 주제 특성상 전 문서에서 공통으로 나타나 추가적인 의미 파악이 불필요할 것으로 판단된 '외교', '한국', '장관' 등 3개 단어는 분석에서 제외하였고, 동시에 연설문이라는 특성상 모든 문서에서 높은 빈도로 나타나 대외정책 분석과 무관할 것으로 평가되는 '귀빈', '여러분', '우리', '오늘', '자리', '내외', '감사', '포럼', '개최', '회의' 등의 단어 역시 선별하여 불용어 처리하였다.

2) 빈도분석(Term Frequency Analysis)

전처리를 거친 주제어를 대상으로 각 문서 등장 빈도를 표현한 행렬(Term-document matrix: TDM)을 작성하고, 문서별로 특정 단어 출현 빈도를 측정하는 방식을 빈도분석이라고 한다. 어떤 문서가 특정 주제를 다룬다면 연관된 단어가 자주 등장하게

되는데, 이를 통해 전체 문서에 대한 요약 정보를 추출할 수 있다(Krippendorff 2004)는 점에서 빈도분석은 텍스트마이닝의 기초이자, 필수 단계로 여겨진다. 물론 단어빈도로만 상대적 중요도를 판단하는 경우, 어디에서나 등장하여 특별한 정보를 담고 있지 않은 단어가 상위권에 위치함으로 전체적인 분석 품질을 떨어뜨릴 수 있다는 우려도 제기되며, 이에 따라 TF-IDF(Term Frequency-Inverse Document Frequency, 단어빈도-역문서빈도)를 활용한 방식이 권장되기도 한다(Ramos 2003). TF-IDF는 과도하게 등장하는 특정 단어가 강조되는 것을 방지하기 위해 몇몇 문서에만 등장하는 단어를 오히려 정보량이 많은 단어로 평가하고 그에 대해 가중치를 부여하는 방식이다[4]. 다만, 해당 분석 대상인 연설문은 매우 정제되고 간결한 언어로 쓰인 점을 고려하고, 짧은 분량의 연설문의 경우에는 해당 글을 구성하는 단어수가 제한되어 문서 내 특정 단어가 오히려 지나치게 과대평가되는 경향을 방지할 필요가 있었다. 따라서 전체 문서를 기준으로 주요 단어를 확인하기 위해 단어빈도를 계산하고, 특징을 한눈에 표현하고자 하였다. 시각화의 방식으로는 워드클라우드도 효과적인 도구이지만, 본 연구에서는 주요 단어의 구체적인 빈도수 및 순위 확인을 위해 상위 20개 단어를 내림차순으로 정렬한 그래프로 특징을 파악하고자 한다.

3) 토픽모델링(Topic Modeling)

토픽분석을 위해서는 문서-단어 행렬(Document-term matrix: DTM)을 생성하여 분석에 활용하는데, 빈도분석과 같은 의도에서 TF-IDF 가중치 행렬 대신, 가중치를 부여하지 않은 Binary 행렬을 활용하였다. 관련해서는 복수의 선행연구(김지은 2017, 임화진·박성현 2015, 이성직·김한준 2009)가 고유명사나 새로운 개념을 분석할 때, 뉴스 등 단문을 분석대상으로 활용할 때 등 TF-IDF를 이용하는 경우 되려 실제보다 단어의 중요도가 과소 및 과대 평가될 수 있는 상황에 대해 주의를 요구한다고 지적하고 있다.

토픽모델링은 각 문서 내 존재하는 복수(k)의 주제가 나타날 확률을 기반으로 특정 문서의 특성을 나타내는 기법인 점을 고려할 때,(Blei and Ng et al 2003) 전체 문서 집합을 가장 효과적으로 표현할 수 있는 숨은 변인(hidden variable)의

4) 본고는 TF-IDF를 직접적으로 분석에 활용하지 않으므로 분석방법 및 관련 이론에 대한 상세한 설명은 참고문헌(Salton and Buckley 1988, Ramos 2003)으로 갈음하고자 한다.

수 k를 미리 지정하는 것이 중요하다. 선행연구에 따르면 반복적인 모델 생성 후, 이 모델들을 통계적으로 비교하여 적정값을 찾거나 혹은 여러 k 값의 결과에 대한 연구자의 직관적 판단으로 값을 찾을 수도 있다(김영욱·함승경 외 2017). 두 가지 방법 모두 장단점이 있으나 본 연구에서는 연구자가 외교백서를 근거로 전체 연설문 제목을 사전 검토한 후, 2~10개 사이에서 반복적으로 k를 지정하고 결과를 비교하는 과정을 거쳐 정부별 최적 토픽 수를 각각 설정하였다.

3. 연구문제

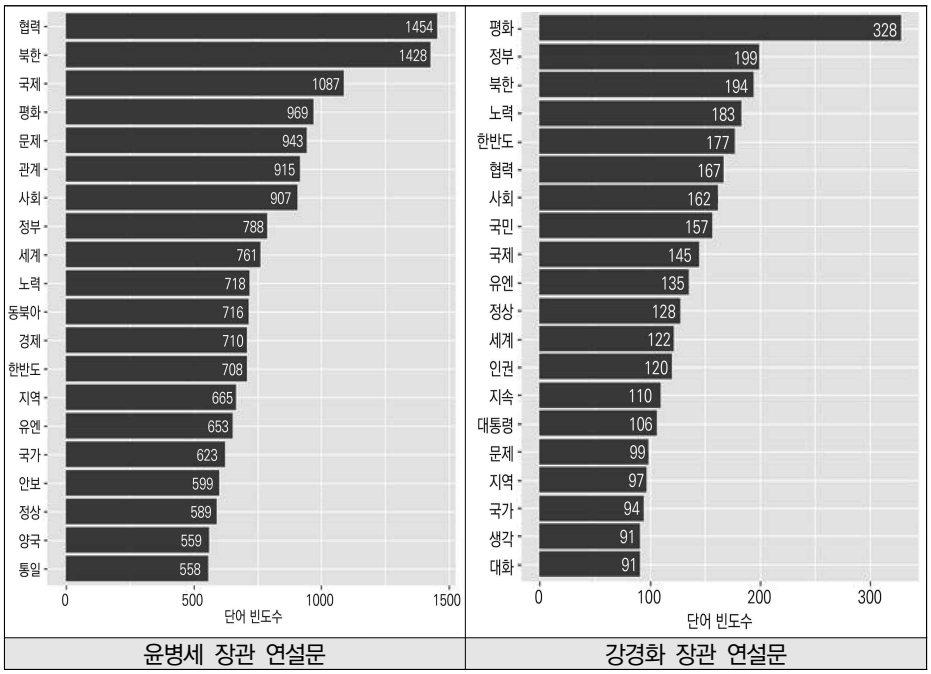
본 연구는 박근혜 정부에서 문재인 정부로 정권이 변화함에 따라 한국의 대외정책에는 어떤 변화가 있었는지를 살펴보고자 한다. 세부적으로 정부별 외교부 장관 연설문 내 빈출단어 및 토픽을 분석하고, 그 결과를 외교백서와 비교해 봄으로써 외교정책 분야에 빅데이터분석 특히 그중에서도 텍스트마이닝 기법을 활용한 연구 분석의 함의를 도출하고자 한다. 한편 동 연구의 목적이 연설문에 기반한 정부별 주요 키워드 및 주제를 찾아내는 것인 만큼 기존 배경 지식에 기반한 해석이나 평가는 가능한 생략 또는 지양하고, 연구를 통해 검증된 사실 중심으로 글을 구성하고자 한다.

V. 분석결과

1. 빈도분석(Term Frequency Analysis)을 활용한 연설문 키워드 분석

정부별 외교부 장관 연설문을 대상으로 텍스트마이닝 분석을 수행한 결과, 각각의 상위 20개 빈출 단어는 다음의 〈그림 2〉와 같다.

〈그림 2〉 정부별 외교부 장관 연설문 내 상위 20개 빈출 키워드



빈도분석 결과에 따르면 윤병세 장관의 연설문에서 최빈출 단어는 '협력'으로 총 1,454회 사용되었으며, 그 밖에 '북한', '국제', '평화', '문제' 등의 단어가 자주 언급되었다. 한편, 강경화 장관의 연설문에서 최다 등장한 단어는 '평화'로 총 328회 언급되었고, 그 밖에 '정부', '북한', '노력', '한반도'가 빈번하게 등장하였다.

정부 간 공통으로 사용된 주요 단어는 대외정책의 방식인 '협력', '노력', 대외정책의 주체 또는 대상으로 '세계', '사회', '지역', '정부', '정상' 등 이었다. 그리고 대외정책의 주요 이슈로는 '북한' 및 '한반도'가 동시에 상위권에 포함되었으며, 대외정책의 목적으로 '평화'가 정부별로 각각 4위 및 1위라는 높은 순위를 기록한 것을 통해 양 정부에서 평화가 외교정책 목표로 높은 의미가 있음을 추론할 수 있다.

반면 특정 장관의 연설문에서만 나타나는 키워드를 통해서는 정부 간 차별화된

대외정책 이슈를 짐작할 수 있다. 윤병세 장관의 연설문에서만 나타난 주요 단어에는 '동북아', '경제', '안보', '통일' 등이 있다. 이와 달리 강경화 장관의 연설문에서는 '국민', '인권', '지속', '대화' 등의 단어가 주요하게 나타나 정권 교체에 따라 대외정책을 드러내는 핵심 단어에도 상당 부분 변화가 발생하였음을 알게 된다.

2. LDA 기반 토픽모델링 분석

1) 토픽모델링 분석 수행

정부 간 주요 외교정책 의제 비교를 위해 추가로 토픽모델링(Topic Modeling)을 통한 중심 주제 및 관련 키워드를 추출하였다. 각 문서가 최대한 고르게 토픽에 분포될 수 있으면서도 대표 문서를 충분히 설명하는 토픽으로 구성된 모델링을 찾기 위해 본 연구에서는 2개에서 10개까지 반복적인 사전 테스트를 통해서 적절한 토픽의 개수를 설정하였다. 그 결과, 박근혜 정부의 주요 토픽은 총 8개로 구성되었고, 각각의 토픽을 구성하는 주요 단어를 아래와 같이 15개씩 빈도순으로 배치하였다. 토픽별 주제는 도출된 주요 단어 및 토픽별 대표 문서 세부 검토를 통한 연구자의 정성적인 판단하에 설정하되, 기존 외교백서 검토 결과를 참조함으로 자의적인 판단 여지를 줄이고자 하였다.

〈표 4〉 윤병세 장관 연설문 토픽모델링 결과

주제 / 순위	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
	한반도 평화 통일과 유라시아 이니셔티브	국민행복 시대를 열어가는 외교	국익을 위한 외교·안보 전략	북한 인권 문제 및 안보 위협에 따른 국제사회 대응	경제 협력 및 교류 확대로 파트너십 확대	국제사회 내 한국의 주요 현안 해결 노력	SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할	동북아 평화협력 구상
1	한반도	국민	관계	북한	협력	문제	개발	협력
2	통일	세계	정상	인권	양국	정부	사회	동북아
3	역사	국가	전략	유엔	경제	국제	유엔	평화
4	평화	행복	생각	국제	지역	사회	지속	신뢰
5	유라시아	문화	대통령	사회	발전	일본	기여	지역

6	독일	시대	글로벌	안보리	분야	중요	가능	아시아
7	국제	노력	미국	위협	관계	해결	국제	국가
8	도전	성장	최근	안보	강화	관련	기후	구상
9	변화	발전	동맹	결의	중동	합의	총회	역내
10	정책	적극	회담	대응	중요	노력	목표	유럽
11	박근혜	나라	협의	미사일	공동	주요	변화	구축
12	대통령	정부	안보	상황	방문	정책	사이버	안보
13	상황	마음	관련	핵실험	미래	이후	세계	세계
14	경제	모두	계기	조치	증진	인식	중요	동아시아
15	분단	대사	차원	채택	대화	포함	노력	비전

토픽모델링을 실시하면 각 주제를 구성하는 연관 단어 묶음을 도출할 수 있고, 이 단어를 조합함으로써 토픽별 핵심주제를 유추할 수 있다. 토픽1은 '한반도', '통일', '평화', '유라시아' 등의 단어를 중심으로 구성되어 있다는 점에서 '한반도 평화통일과 유라시아 이니셔티브'라는 주제로 귀결된다. 박근혜 대통령이 제안한 유라시아 이니셔티브는 세계 최대 단일 대륙이자 거대 시장인 유라시아 역내 국가 간 협력을 도모하고, 북한의 개방을 유도해 한반도의 긴장을 완화하고 통일 기반을 구축한다는 내용을 골자로 한다.

토픽 2에서 나타난 주요 단어로는 '국민', '행복', '시대' 등을 들 수 있으며, 외교의 목적을 국민을 행복하게 하는 외교, 국민 중심의 외교를 주장한 외교부 정책 기조와 일치한다. 따라서 토픽 2 주제도 '국민행복시대를 열어가는 외교'로 결정하였다.

'관계', '정상', '회담', '전략' 등의 단어가 상위에 나타난 토픽 3은 연설문별로 다루는 세부주제가 다양하고 포괄적이었다. 그러나 공통적으로 정상회담 등 마일스톤을 계기로 국익 증진을 위한 한국의 외교전략을 조망하는 내용이 포함되었다. 이에 따라 토픽 3의 주제는 포괄적인 의미로 '국익을 위한 외교·안보 전략'으로 설정하였다.

토픽 4의 경우, 토픽 1과 마찬가지로 북한 이슈를 다룬다는 공통점이 있다. 그러나 토픽 1이 한반도 안정에 대한 미래지향적 기대를 반영한다면, 토픽 4에서는 현재 당면한 북한 관련 문제인 '인권', '미사일', '핵실험' 등의 단어가 중심에 자리한다. 한편, 문제 해결 방식과 관련된 단어로 '유엔', '안보리', '결의', '채택' 등이 활용되고 있다는 점에서 북핵문제를 남북 양자 간의 문제로 국한해 다루기보다,

다자협력의 틀에서 해법을 찾고자 하였음을 알 수 있다. 결론적으로 토픽 4의 주제는 '북한 인권문제 및 안보 위협에 따른 국제사회 대응'으로 요약될 수 있다.

다음으로 토픽 5에서 다루어진 주제는 '경제 협력 및 교류 확대로 파트너십 확대'이다. '경제' 및 '협력'이 핵심 단어로 등장하였으며, 주요 경제 파트너였던 '중동'이 높은 빈도로 해당 토픽에 등장하였다. 한편, 관계를 의미하는 단어로 '공동', '방문', '대화' 등의 어휘가 함께 쓰이고 있었다.

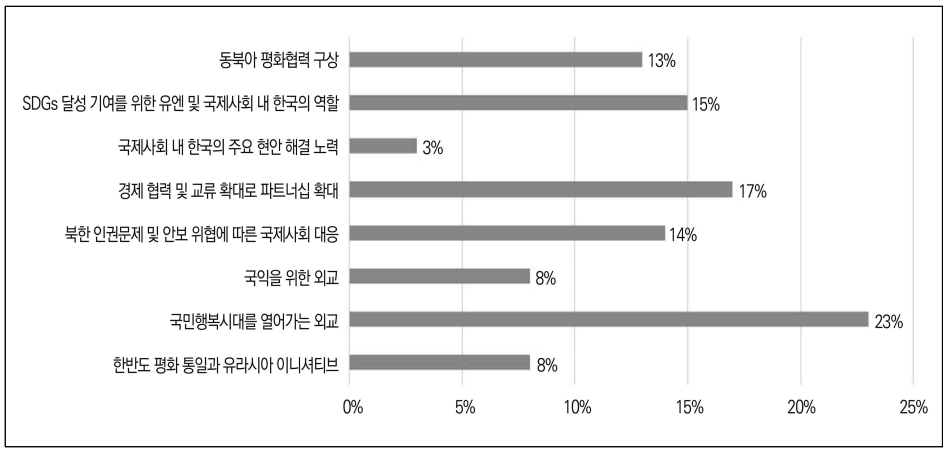
이어진 주제에서는 박근혜 정부에서 중요하게 다룬 위안부 피해자 문제 해결을 위한 한일 양국 합의 이슈 등이 포함되었다. 그 밖에도 해당 토픽에서는 한국과 관련된 여러 국제 현안을 원만히 해결하고자 하는 외교적 노력이 강조되고 있다. 주요 대상국으로 '일본'이 언급되었으며, '문제', '해결' 그리고 해결 방안으로 '합의', '노력' 등의 단어가 등장하였다.

토픽 7에서는 UN에서 주창한 지속 가능한 개발목표(SDGs: Sustainable Development Goals)를 중심으로 글로벌 이슈 해결을 위한 국제사회 공동 노력에 기여하고자 하는 한국의 의지를 다루고 있다. 세부 분야로는 '기후', '사이버' 등이 강조되고 있었다.

마지막으로 토픽 8번에서는 '동북아', '평화', '신뢰', '안보' '협력' 등의 단어가 자주 언급되었다. 동북아평화협력 구상 및 해당 이니셔티브의 협업 대상 지역으로 (동)아시아 및 유럽 지역이 함께 등장한다.

박근혜 정부 시기 등장한 연설문을 토픽별로 재분류한 결과, 토픽 2(국민행복시대를 열어가는 외교)가 등장할 확률이 가장 높게 나타난 연설문이 총 49건으로 전체의 4분의 1(23%)을 차지하였으며, 토픽 5(경제 협력 및 교류 확대로 파트너십 확대)가 17%로 2위, 토픽 7(SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할)이 15%로 3위, 토픽 4(북한 인권문제 및 안보 위협에 따른 국제사회 대응)가 근소한 차이로 4위를 각각 차지하였다. 그 밖에 토픽 8(동북아 평화협력 구상)이 5위(13%), 토픽 1(한반도 평화통일과 유라시아 이니셔티브)이 6위(8%), 토픽 3(국익을 위한 외교·안보 전략)이 7위(8%), 끝으로 토픽 6(국제사회 내 한국의 주요 현안 해결 노력)이 7%로 가장 낮은 순위를 차지하였다.

〈그림 3〉 윤병세 장관 연설문 토픽별 비중



이어서 문재인 정부의 장관 연설문에 대한 토픽모델링 분석을 통해 추출된 주제는 총 6개이며, 이를 구성하는 상위 15개 단어 및 연구자가 설정한 토픽별 주제는 아래 표와 같다.

〈표 5〉 강경화 장관 연설문 토픽모델링 결과

주제 / 순위	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
	평창올림픽 계기 대화를 통한 북핵 문제 해결	국가 및 지역 간 경제 협력 증진을 통한 발전 모색	SDGs 달성을 위한 국제사회 지원 노력	한반도 및 세계평화를 위한 남북 및 북미 간 정상회담	유엔 내 한국의 여성·평화·인권문제 기여 노력	국민과 소통하는 외교
1	북한	협력	국제	평화	유엔	국민
2	대통령	지역	지속	한반도	인권	나라
3	대화	국가	중요	세계	정부	정책
4	정부	경제	변화	정상	평화	발전
5	문제	생각	강화	회담	사회	국제
6	비핵화	관계	안보	방문	지원	소통
7	올림픽	필요	개발	남북	노력	생각
8	양국	분야	지지	번영	여성	과제
9	공동	기회	노력	역사	대한민국	적극
10	정상	대사	논의	노력	상황	노력

11	해결	성공	진전	완전	기여	역할
12	평창	최근	평가	북미	과정	추진
13	제재	출범	이행	시간	유지	기대
14	미사일	참석	달성	합의	도전	사회
15	미국	동북아	대응	체제	대표	직원

토픽별 핵심주제를 유추하기 위해 토픽모델링 결과에 따른 우선순위 단어를 살펴본 결과, 우선 토픽 1은 '북한', '대화', '비핵화', '평창', '올림픽' 등의 단어를 중심으로 구성되어 있었다. 이를 통해 2018년 평창 동계올림픽 개최를 계기로 북한과 화해 분위기를 조성하고, 북핵 문제를 대화를 통해 해결하겠다는 문재인 정부의 의지를 확인할 수 있었다.

토픽 2에서 나타난 주요 단어에는 '협력', '지역', '국가', '경제' 등이 있다. 이를 통해 다양한 국가 및 지역과의 상호 경제 협력 증진을 통해 공동 발전을 모색하고자 하는 한국 정부의 정책 과제를 발견하게 되었다.

'국제', '지속', '개발' 등의 단어가 상단에 위치한 토픽 3의 경우는 국제사회의 대표적인 공동 과제로 인식되는 SDGs를 중심으로 국제사회의 이행 노력과 목표 달성에 대한 평가, 관련 논의의 진전 등에 대해 다루고 있었다. 또한, 연설문을 통해 한국 정부 역시 해당 목표 달성에 높은 관심을 가지고 적극적으로 대응하고 있음이 강조되었다.

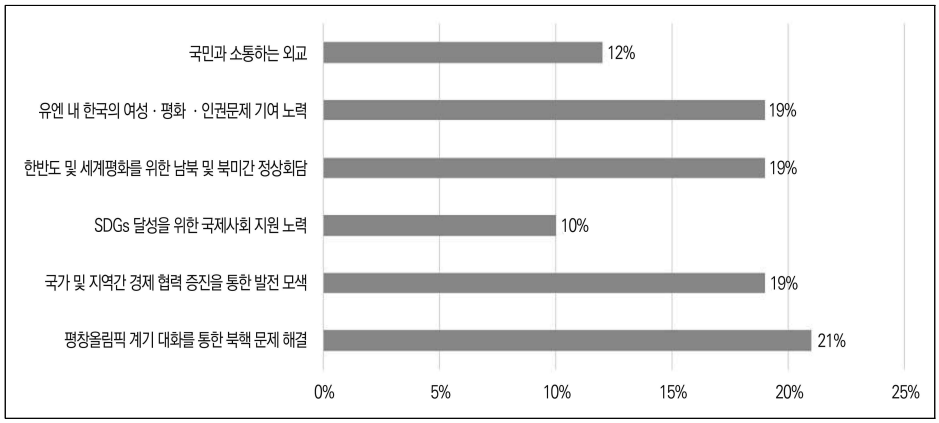
토픽 4의 경우, 북한, 미국 등 토픽 1과 일부 유사한 단어가 있지만, 초점이 2018년~2019년 사이 이루어진 1·2차 남북정상회담 및 북미정상회담에 맞추어져 있다는 점에서 차별화된다. 또한, 한반도 및 세계평화 정착, 공동번영을 이루겠다는 한국 정부의 기대가 전면에 드러난 주제라는 점에서 별도의 토픽으로 다루는 것이 적절하다고 판단되었다.

다음으로 토픽 5에서 다루어진 주제는 UN을 필두로 한 글로벌 다자협력체제 내에서 한국의 역할에 대해 언급하고 있다. 대표적으로 '여성', '평화', '인권' 등 범세계적인 이슈 해결에 적극적으로 이바지하고자 하는 정부의 노력과 의지를 확인할 수 있다.

1~5번까지의 주제가 대외정책의 주요 과제였다면, 마지막 토픽 6은 국내를 대상으로 '국민', '(외교부) 직원' 등이 주로 언급되고 있다. 한편, 주로 등장하는 '기대',

‘역할’, ‘소통’ 등의 단어를 종합적으로 판단할 때, 국민과의 적극적인 소통 노력을 강조하는 국민 중심 외교가 정부의 또 다른 핵심 과제임을 판단할 수 있었다.

〈그림 4〉 강경화 장관 연설문 토픽별 비중



문재인 정부의 외교부 장관 연설문에서 가장 빈번하게 등장한 주제는 토픽 1(평창 올림픽 계기 대화를 통한 북핵 문제 해결)로 전체의 21%를 차지하였다. 이어서 토픽 2(국가 및 지역 간 경제 협력 증진을 통한 발전 모색), 토픽 4(한반도 및 세계평화를 위한 남북 및 북미 간 정상회담), 토픽 5(유엔 내 한국의 여성·평화·인권문제 기여 노력) 등 3개의 주제가 모두 19%로 고르게 자주 등장함을 알 수 있었다. 끝으로 토픽 6(국민과 소통하는 외교) 및 토픽 3(SDGs 달성을 위한 국제사회 지원 노력)이 각각 12%, 10%로 나타나 상대적으로 낮은 빈도로 등장하였다.

2) 토픽모델링 결과 비교

각각의 토픽모델링 결과를 사전 검토한 외교백서 내 대외정책 과제와 비교하면, 명시적으로 언급된 공식 문서와 연설문으로 나타난 외교정책 과제 간 공통점과 차이점을 살펴볼 수 있을 것이다. 우선 아래는 2014~2016년 외교백서와 윤병세 장관 연설문 토픽모델링 결과를 비교한 표이다. 주제는 외교백서에서 나타난 대외정책 주제를 기초로 하되 백서 내 언급되지 않았으나, 연설문상 토픽으로 분류된 경우

에는 해당 내용을 추가하였다.

〈표 6〉 2014~2016년 외교백서와 윤병세 장관 연설문 토픽모델링 결과 비교

주제 \ 정부		외교백서	연설문
1	북한	북핵문제 진전 및 평화통일	북한 인권문제 및 안보 위협에 따른 국제사회 대응
2	지역	동북아 평화협력 구상과 유라시아 협력 확대	한반도 평화통일과 유라시아 이니셔티브
			동북아 평화협력 구상
3	주변국	전략적 로드맵에 따른 주변국 외교	-
4	국익	-	국익을 위한 외교·안보 전략
5	국제사회	세계평화와 발전에 기여하는 책임 있는 중견국 실현	SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할
			국제사회 내 한국의 주요 현안 해결 노력
6	경제	경제외교를 통한 경제부흥 선도	경제 협력 및 교류 확대로 파트너십 확대
7	국민	재외국민 안전·권익 보호와 공공외교 확대	국민행복시대를 열어가는 외교

넓은 관점에서의 주제는 외교백서 상 대외정책 과제와 크게 다르지 않은 것을 알 수 있다. 다만 하나씩 차이점을 짚어보자면, 외교백서에서 북한을 북핵문제와 평화통일의 두가지 주제를 포괄적으로 다룬 것과 달리 실제 장관 연설문을 통해서는 북핵문제를 포함한 안보 위협과 함께 북한의 인권문제를 중심으로 대북정책에 접근하였으며, 외교백서에서는 남북 양자관계를 위주로 현안을 다루고 있는 것처럼 보였으나, 연설문을 통해 실제로는 박근혜 정부가 안보리를 비롯한 국제사회 속에서 해결점을 모색하였음을 알게 되었다. 또한, 통일이라는 남북관계의 가장 큰 숙제 또한 유라시아 이니셔티브라고 하는 지역 협업체계를 기초로 접근하였으며, 평화라는 테마에서 유사성을 보이는 동북아 평화협력과 유라시아 이니셔티브는 사실 별도의 주제로 다루어지고 있었다. 다음으로 주변 4강에 대해서는 외교백서와 달리 실제 연설문 속에 독립적인 토픽으로 드러나지 않았다. 다만, 미국의 경우 국익을 위한 외교·안보전략 차원에서 높은 순위에 언급되었고, 러시아는 유라시아 이니셔티브에서, 중국은 동북아 평화협력구상에서 간접적으로 나타났으며, 일본은 국제사회 내

한국의 주요 현안 해결 노력에서 살펴볼 수 있었는데, 특히 위안부 피해자 문제가 주요 이슈였음이 핵심 주제어를 통해 드러난다. 다음으로 국제사회 관련 아젠다는 세계평화와 발전이라는 다소 모호한 주제가 연설문을 통해 기후 및 사이버보안 등의 이니셔티브를 통해 SDGs에 한국이 기여하는 방향으로 구체화되었고, 위안부 이슈 역시 한·일 문제로 국한시키기 보다 이것을 국제사회 과제로 부각하려는 노력이 확인되었다. 한편, 경제적 지향점과 관련하여 외교백서와 연설문 간 뚜렷한 차이는 없으나 연설문을 통해 중동지역이 주요 대상인 점과 접근 방식으로 협력, 교류 확대 등 상호호혜적인 방향을 모색한 점 등을 발견하였다. 끝으로 연설문 속에 나타난 국민 관련 주제는 외교의 궁극적인 지향점이 국민을 행복하게 하는 것이라는 다소 추상적인 내용으로 공공외교나 재외국민 보호와 관련된 구체적인 활동보다는 구호나 선언에 가까운 표현이 다수 등장하였다.

이어서 2017~2019년 외교백서와 강경화 장관 연설문 토픽모델링 결과를 주요 정책과제 중심으로 비교한 표는 아래와 같다.

〈표 7〉 2017~2019년 외교백서와 강경화 장관 연설문 토픽모델링 결과 비교

주제 \ 정부		외교백서	연설문
1	북한	한반도 평화프로세스 진전	평창올림픽 계기 대화를 통한 북핵 문제 해결
			한반도 및 세계평화를 위한 남북 및 북미 간 정상회담
2	주변국	주변 4국과 협력외교 강화	-
3	지역	외교 다변화를 통한 외교 지평 확대	국가 및 지역 간 경제 협력 증진을 통한 발전 모색
4	국제사회	국제사회 내 역할 확대와 위상 제고	SDGs 달성을 위한 국제사회 지원 노력
			유엔 내 한국의 여성·평화·인권문제 기여 노력
5	국민	국민과 함께하는 외교 실현	국민과 소통하는 외교

전반적으로 문재인 정부 외교백서 상에 제시된 대외정책 목표는 연설문을 통해 뚜렷하게 구체화되고 있었다. 한반도 평화프로세스 진전에 해당하는 연설문 상 토픽은 평창올림픽 계기 대화를 통한 북핵 문제 해결과 한반도 및 세계평화를 위한 남북 및 북미 정상회담이라는 두 축으로 구성되어 있으며, 이는 주변국 중 한미 관계와도 연결되어 있다. 다만, 미국을 제외하면 특별한 주변국 관련 언급은 찾아보기 힘들었고 다양한 지역과의 협력, 그중에서도 특히 경제 협력 증진에 초점이 맞춰져 있다는 점이 새롭게 발견된 점이었다. 다음 국제사회와의 관계에서는 SDGs 달성 그리고 특별히 여성, 평화, 인권문제 등에 중점을 두고 있음이 확인되었다. 그러나 대국민 정책으로 소통과 참여를 강조한다는 점에서는 외교백서와의 큰 차이를 발견하기 어려웠다.

박근혜 정부와 문재인 정부 간 토픽별 중심 내용이 완전히 일치하지는 않으나, 유사 주제별로 분류해 비교한 결과는 아래 표와 같다.

〈표 8〉 정부별 연설문 주제 비교

주제 \ 정부		박근혜 정부	문재인 정부
1	북한	북한 인권문제 및 안보 위협에 따른 국제사회 대응	평창올림픽 계기 대화를 통한 북핵 문제 해결
2	한반도	한반도 평화통일과 유라시아 이니셔티브	한반도 및 세계평화를 위한 남북 및 북미 간 정상회담
3	동북아	동북아 평화협력 구상	-
4	국익	국익을 위한 외교·안보 전략	-
5	국제사회	SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할	SDGs 달성을 위한 국제사회 지원 노력
		국제사회 내 한국의 주요 현안 해결 노력	유엔 내 한국의 여성·평화·인권문제 기여 노력
6	경제	경제 협력 및 교류 확대로 파트너십 확대	국가 및 지역 간 경제 협력 증진을 통한 발전 모색
7	국민	국민행복시대를 열어가는 외교	국민과 소통하는 외교

박근혜 정부의 북한 관련 토픽이 인권, 안보 위협 중심으로 구성되었다면, 문재인 정부는 평창올림픽 계기 대화를 통한 북핵 문제 해결이라는 점에서 뚜렷한 차이를 보인다. 한반도 관련해서는 평화통일, 유라시아 이니셔티브가 박근혜 정부의 주요 정책이라면, 문재인 정부는 한반도 및 세계평화를 목표로 남·북·미 간 정상회담이 주요 현안임을 알 수 있다. 한편, 박근혜 정부에서만 나타난 토픽으로는 동북아 평화협력구상 및 국익을 위한 양·다자 외교 및 안보전략이 있었고 국제사회 관련 토픽으로 SDG 달성은 양 정부가 공유하는 공통 목표임을 알 수 있었다. 다만 세부 토픽을 통해 정부의 주된 협력 관심사가 기후변화 및 사이버보안에서 여성, 평화, 인권 등 거버넌스 관련 주제로 옮겨갔음이 확인되었다. 한편, 박근혜 정부는 추가로 한일관계 회복 등 국내외 정치가 함께 연관된 이슈를 대외정책 주제 중 하나로 다루고 있었다. 경제 이슈 역시 양측에서 모두 중요하게 다루어지고 있었으며, 끝으로 국민은 주요한 대외정책 토픽 중 하나였지만 핵심 키워드는 '행복'과 '소통'으로 대외정책 추진 과정에서 국민을 바라보는 관점에 차이가 있음을 확인하였다.

3) 토픽별 점유율 및 시계열 변화 추적

수집한 연설문 전체에 대한 토픽모델링을 실시할 경우, 정부별 중점과제를 한눈에 파악할 수 있다는 장점이 있으나, 시간의 흐름에 따라 한 정부 내에서도 특정 주제에 대한 관심도가 변화할 수 있다는 점에서 해당 내용을 파악하기 어려운 한계가 있다. 그러나 토픽을 시계열적으로 표현하면 연설문상 토픽의 상대적 비중 변화를 통해 특정 시기에 각 정부에서 중요하게 다루어진 대외정책 주제는 무엇인지, 우선순위가 어떻게 바뀌고 있는지를 면밀하게 살펴볼 수 있다. 이를 위해 아래 표에서는 윤병세 장관 재임 기간을 분기별로 구분하고, 연설문에 나타난 토픽 각각의 등장 비율을 순위로 환산하여 1~8위까지로 표현하였다. 더불어 우선순위가 높을수록 짙은 음영을 주어, 시기별 주요 토픽을 한눈에 알아볼 수 있도록 표시하였다.

〈표 9〉 윤병세 장관 재임기간(2013~2017) 연설문 토픽 순위 분석

주제 / 연도		Topic 1 한반도 평화 통일과 유라시아 이니셔티브	Topic 2 국민행복 시대를 열어가는 외교	Topic 3 국익을 위한 외교·안보 전략	Topic 4 북한 인권 문제 및 안보 위협에 따른 국제사회 대응	Topic 5 경제 협력 및 교류 확대로 파트너십 확대	Topic 6 국제사회 내 한국의 주요 현안 해결 노력	Topic 7 SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할	Topic 8 동북아 평화협력 구상
2013	2	3	1	6	8	4	7	5	2
	3	8	1	5	6	4	7	2	3
	4	7	2	4	8	6	5	1	3
2014	1	2	1	6	7	4	3	8	5
	2	1	2	7	6	8	5	4	3
	3	6	3	4	8	1	5	7	2
	4	3	5	4	8	1	6	7	2
2015	1	2	5	4	8	1	6	7	3
	2	8	6	3	7	4	5	1	2
	3	2	4	3	8	5	6	1	7
	4	7	6	2	8	4	1	3	5
2016	1	7	5	3	1	6	2	4	8
	2	6	7	3	2	1	4	5	8
	3	8	5	4	1	3	6	2	7
	4	5	8	2	1	6	3	4	7
2017	1	6	3	8	1	7	5	2	4
	2	5	2	3	1	4	6	7	8

토픽 1(한반도 평화 통일과 유라시아 이니셔티브)은 상대적으로 2013년 2분기, 2014년~2015년 3분기까지 집중적으로 나타나는 주제였고, 2014년도 2분기 외교 정책의 가장 큰 화두로 기록되었다. 토픽 2(국민행복시대를 열어가는 외교)는 정부 초반(2013년~2014년) 지속적인 상위권을 기록하였으나, 2016년 4분기까지 점차 비중이 줄어들어 언급이 많지 않았고, 2017년 정부 말기에 다시 등장하는 양상을 보였다. 토픽 3(국익을 위한 외교·안보 전략)은 한 번도 주요 토픽 중 1위를 기록한 적은 없으나, 2015년 2분기에서 2016년 4분기 사이 꾸준히 상위권을 기록하였다. 반면, 토픽 4(북한 인권문제 및 안보 위협에 따른 국제사회 대응)는 정부 초반에는 거의 언급되지 않던 주제였다. 그러나 2016년 1분기 이후, 2분기를 제외한 전 기간 1순위를 차지할 만큼 박근혜 정부의 주요 관심사로 자리매김하였다. 토픽 5(경제 협력 및 교류 확대로 파트너십 확대)는 정권 중반부, 즉 2014년 3분기에서 2016년 2분기까지 부침은 있지만 꾸준하게 등장하고 있으며 특히, 2014년 2분기~2015년

1분기에 연속 1위 점유율을 보였다는 점에서 경제외교의 우선순위가 높았던 시점을 확인할 수 있다. 토픽 6(국제사회 내 한국의 주요 현안 해결 노력)은 전반적으로 전면에 드러나는 주제는 아니었는데, 2015년 4분기에 1위를 기록하면서 갑작스럽게 중심 화두로 등장하였고, 이후 2016년 4분기 사이에 주로 언급되었다. 토픽 7(SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할)은 2013년 하반기에 집중적으로 언급되다가 2014년도에는 관심사에서 사라진 것처럼 보였다. 그러나 2015년 2분기에 다시 급등장 하여 정권 말기까지 꾸준하게 비중을 차지하였다. 끝으로 토픽 8(동북아 평화협력 구상)의 경우, 1위를 차지한 적은 없으나, 매우 뚜렷하게 박근혜 정부 출범 초중반(2013년~2015년)에 주로 나타나다가 2016년 이후에는 거의 관심사에서 사라진 것처럼 보인다.

마찬가지 방식으로 아래는 문재인 정부에서 중요하게 다루어진 연설문 주제의 변화 양상을 순위로 나타낸 표이다.

〈표 10〉 강경화 장관 재임기간(2017~2019) 연설문 토픽 순위 분석

주제 / 연도		Topic 1 평창올림픽 계기 대화를 통한 북핵 문제 해결	Topic 2 국가 및 지역 간 경제 협력 증진을 통한 발전 모색	Topic 3 SDGs 달성을 위한 국제사회 지원 노력	Topic 4 한반도 및 세계평화를 위한 남북 및 북미 간 정상회담	Topic 5 유엔 내 한국의 여성·평화·인권 문제 기여 노력	Topic 6 국민과 소통하는 외교
2017	2	1	5	6	3	4	2
	3	1	2	3	4	6	5
	4	3	1	2	6	5	4
2018	1	1	3	2	5	4	6
	2	4	2	6	1	3	5
	3	5	2	4	3	6	1
	4	3	5	6	2	4	1
2019	1	5	4	2	3	1	6
	2	6	3	5	2	1	4
	3	6	4	2	5	1	3
	4	5	3	2	4	1	6
2020	1	3	4	1	5	2	6

토픽 1(평창올림픽 계기 대화를 통한 북핵 문제 해결)은 2017~2018년 1분기까지 2017년 4분기를 제외하고는 지속 1위를 기록한 대외정책 화두였다. 다만 2018년 2분기 이후로는 전반적으로 확연히 등장 빈도가 감소하였다. 다음으로 토픽 2(국가 및 지역 간 경제 협력 증진을 통한 발전 모색)는 2017년 3분기부터 2018년 3분기 사이 1년간 꾸준하게 등장하다 이후에는 다소 비중이 줄어드는 모습이 포착되었고, 토픽 3(SDGs 달성을 위한 국제사회 지원 노력)은 특정 시기마다 등장하는데 2017년 4분기~2018년 1분기 사이 및 2019년 이후 매우 중요한 위치를 차지함을 알게 되었다. 한편, 토픽 4(한반도 및 세계평화를 위한 남북 및 북미 간 정상회담)는 2018년 2분기~2019년 2분기 1년간 상위권을 유지하였으나, 2019년 3분기 이후 급격히 등장 빈도가 감소하였다. 다음 토픽 5(유엔 내 한국의 여성·평화·인권 문제 기여 노력)는 토픽 3와 함께 국제개발협력 이슈로서 2019년 이후 문재인 정부 외교정책의 가장 큰 관심사임을 알 수 있었다. 끝으로 토픽 6(국민과 소통하는 외교)은 2018년 3~4분기에 빈번하게 등장하였고, 2019년 이후에는 거의 언급되지 않는 토픽임이 확인되었다.

Ⅵ. 결론 및 시사점

본 연구에서는 박근혜 정부에서 문재인 정부로 정권이 변화함에 따라, 한국 대외정책에는 어떤 변화가 있었는지를 살펴보고자 키워드 분석 및 토픽모델링 분석을 수행하였다. 연구를 위해 외교부 웹사이트에서 2013년에서 2020년 사이 총 255건의 외교부 장관 연설문을 수집하였고, 분석에는 오픈소프트웨어 프로그램인 R이 활용되었다.

연구를 통해 밝혀진 주요 결과는 다음과 같다. 먼저 양 정부는 공통적으로 북한, 한반도, 평화 등의 키워드를 자주 사용하며, 박근혜 정부에서 특징적으로 나타난 단어는 동북아, 경제, 안보, 통일 등이고, 문재인 정부에서는 국민, 인권, 지속, 대화 등의 단어 빈도가 높게 나타났다. 다음 토픽모델링 분석을 통해 박근혜 정부의 8개(① 한반도 평화통일과 유라시아 이니셔티브, ② 국민행복시대를 열어가는 외교, ③ 국익을 위한 외교·안보 전략, ④ 북한 인권문제 및 안보 위협에 따른 국제사회

대응, ⑤ 경제 협력 및 교류 확대로 파트너십 확대, ⑥ 국제사회 내 한국의 주요 현안 해결 노력, ⑦ SDGs 달성 기여를 위한 유엔 및 국제사회 내 한국의 역할, ⑧ 동북아 평화협력 구상) 주요 토픽을 발견하였고, 문재인 정부에서는 6개(① 평창 올림픽 계기 대화를 통한 북핵 문제 해결, ② 국가 및 지역 간 경제 협력 증진을 통한 발전 모색, ③ SDGs 달성을 위한 국제사회 지원 노력, ④ 한반도 및 세계평화를 위한 남북 및 북미 간 정상회담, ⑤ 유엔 내 한국의 여성·평화·인권 문제 기여 노력, ⑥ 국민과 소통하는 외교)의 대표적 대외정책 이슈를 도출하였다. 또한, 시간의 흐름에 따라 외교 이슈 간 순위의 변화를 파악함으로써 특정 정부 내에서도 시기별 주요 외교정책에 차이가 있음을 실증적으로 확인하였다.

본 연구에서는 상기 결과를 바탕으로 다음과 같은 몇 가지 정책적 시사점을 제시하고자 한다. 첫째, 외교정책에도 텍스트마이닝 등 다양한 빅데이터 기술을 접목한 연구가 활성화되어야 한다. 선행연구에서 나타난 바와 같이 국제정치 및 외교영역에서 빅데이터 기반 연구에 대한 필요성은 반복적으로 제시되었으나, 정작 실제 국내 외교정책에서 이를 적용한 사례는 희박하다. 특히, LDA방식의 토픽모델링은 정량·정성 분석의 강점을 모두 가지면서도 동 연구와 같은 기존 외교정책 분석만 아니라, 신규 정책의제 발굴 등 측면에도 이바지할 수 있다. 둘째, 장관 담화문 분석은 외교백서와 마찬가지로 정부의 공식적인 입장을 대내외에 알리는 역할을 하면서도 시기별 중요하게 다루어지는 주제나 안건을 다양하게 포괄할 수 있다는 점에서 외교백서 등 정부발간 공식자료 분석의 보완재로 충분히 활용 가치가 있다. 마지막으로 외교정책의 세부 과제에 대한 정부별, 시기별 정책 우선순위를 설정하고, 명시된 우선순위에 따른 정책 운용 평가 등에 동 연구가 기여할 수 있을 것이라 기대된다. 동 연구 내 정부별 토픽 및 토픽별 비중을 통해 시기별로 중요하게 다루어지는 정책 이슈를 정량적으로 확인 가능한바, 이는 주요 외교 과제를 사후에 찾아내는데 유용할 뿐 아니라, 중요성에도 불구 상대적으로 덜 다루어진 과제에 대한 정책 우선순위를 새롭게 부여한다는 측면에서도 도움이 될 것이다. 끝으로 북한에 대한 중장기적 관점에서의 외교정책 수립이 필요하다. 동 연구결과에 따르면 외교정책 과제 중 북한은 가장 큰 비중을 차지하는 주요 이슈이면서도 정권의 정치적 이념에 따라 급격한 정책 노선 변화를 보였다. 물론 정권 및 대내외 환경 변화에 따라 “대화”와 “제재”간 우선순위 차이 발생은 있을 수 있으나, 정책 일관성 및 안정성에 부정적인

영향을 미치지 않도록 장기적 관점에서 다양한 정책 수단을 종합적으로 활용하는 방향에서 대북정책이 수립되어야 할 것이다.

마지막으로 동 연구를 진행하면서 나타난 한계점 및 후속 연구를 위한 과제는 다음과 같다. 이번 연구의 목적은 텍스트마이닝 기법을 활용해 박근혜 정부와 문재인 정부 간 주요 대외정책토픽 발견 및 비교에 집중한 대신 각각의 토픽이 지닌 함의 및 도출된 연구결과에 대한 추가적인 질적 분석은 본고에서 거의 다루지 않았다. 이에 따라 배경 지식을 바탕으로 한 동 연구에 대한 정책적 함의는 구체적으로 제시되지 못한 점이 동 연구의 한계로 지적될 수 있다. 다만, 이번 연구를 발판삼아 향후 세부 연구 주제에 대한 연구자의 통찰 및 정성적 관찰이 더해진다면 국내 대외정책 분야 빅데이터 기반 연구가 한층 발전할 수 있으리라 기대해 본다.

투 고 일: 2020. 07. 13.
심사완료일: 2020. 08. 24.
게 재 일: 2020. 11. 30.

참고문헌

권기석, 박진수, 구찬동. 2014.『교육 빅데이터를 활용한 아젠다 개발-사회 연결망 분석을 중심으로』. 서울: 한국교육학술정보원.

그레이엄 앨리슨, 필립 젤리코 지음, 김태현 옮김. 2018.『결정의 본질』. 서울: 모던아카이브

김미숙, 전상완. 2019. “빅데이터 텍스트마이닝 분석을 통한 스포츠정책 변화 추이: 문화체육관광부 장관 취임사를 중심으로”.『한국스포츠학회지』. 제17권 2호.

김상배. 2017. “외교안보 분야 빅데이터 분석: 어떻게 수행할 것인가”.『국제문제연구소 워킹페이퍼』. 제25호.

김영욱, 함승경, 김영지, 최지명. 2017. “사회 쟁점에 대한 비판적 담론 분석”.『커뮤니케이션이론』. 제13권 4호.

김일수. 2015. “외교정책연구의 이론화”.『사회과학연구』. 제32권 2호.

김정미, 윤미영. 2012.『영유아 보육정책에 대한 국민정서분석』. 서울: 한국정보화진흥원.

김지은. 2017. “빅데이터를 활용한 대학구조개혁 평가의 키워드 및 토픽 분석”. 서울대학교 교육학 박사 학위 논문.

김혁. 2016. “대통령의 정책의제설정 과정에서의 정치 커뮤니케이션과 리더십에 관한 연구”.『사이버커뮤니케이션학보』. 제33권 2호.

김형지, 김선우, 김성태. 2018. “스마트폰 중독에 관한 미디어 보도 토픽 모델링 분석”.『방송통신연구』. 제104호.

백영민, 최문호, 장지연. 2014. “한미 정권교체에 따른 주한 미대사관 외교문서의 주제와 감정 표현 변화: 위키리크스 공개 외교전문의 컴퓨터 언어처리 분석”.『언론정보연구』. 제51권 1호.

백우열. 2017. “한국 공공외교 전략 및 정책 연구: 기초적 유형의 개념화”.『국가전략』. 제23권 3호.

서호준. 2019. “우리나라 국방정책의 핵심이슈 도출『2018 국방백서』에 대한 텍스트 네트워크 분석의 적용”.『한국군사』. 제2019-6호.

송태민. 2013. “소셜 빅 데이터 분석을 통한 자살 검색 예측모형 개발”.『보건복지포럼』. 제202호.

외교부. 2014.『2013년 외교백서』.

_____. 2015.『2014년 외교백서』.

_____. 2016.『2015년 외교백서』.
_____. 2017.『2016년 외교백서』.
_____. 2018.『2017년 외교백서』.
_____. 2019.『2018년 외교백서』.
_____. 2020.『2019년 외교백서』.
이수진, 이신화. 2019. "외교정책결정자의 개인적 특성과 미국의 군사적 개입격차: 오바마 대통령의 크림반도 및 남중국해 대응 비교".『국제관계연구』. 제24권 1호.
이창길. 2010. "정권 초기의 가치지향과 정책우선순위: 참여정부와 이명박정부의 언어네트워크 비교분석".『한국행정학보』. 제44권 3호.
임화진, 박성현. 2015. "빅데이터를 이용한 지역미래전략 수립에 관한 시론적 연구".『한국지적정보학회지』. 제17권 1호,
전영환. 2014. "연설문 텍스트마이닝을 통한 미국 연방공개시장위원회 인사들의 통화정책 성향 분석".『대한산업공학회 춘계공동학술대회 논문집』. 서울: 대한산업공학회.
정예림, 강정은. 2019. "기후변화 정책 수립 지원을 위한 소셜 빅데이터 분석".『환경정책』. 제27권 4호,
정원준. 2018. "사드 (THAAD) 이슈를 둘러싼 한국과 중국 간 갈등 쟁점의 변화 추이 연구: 빅데이터를 이용한 시계열 토픽 모델링과 언어 네트워크 분석 기법으로".『한국광고홍보학보』. 제20권 3호
조명현. 2000. "외교정책결정 모형에 관한 연구".『사회과학연구』. 11권.
한국정보화진흥원. 2012. "소셜 분석으로 살펴본 청소년 자살예방정책의 시사점".『IT & Future Strategy』. 제 17호.
홍영희. 2019. "소프트웨어(SW) 교육정책에 대한 토픽분석: 부산 지역 종합지를 중심으로".『통계연구』. 제24권 2호.

Blei, David M., Ng, Andrew Y., and Jordan, Micheal I. 2003. "Latent dirichlet allocation." *Journal of Machine Learning Research*, No. 3: pp. 993-1022.
Eshbaugh-Soha, Matthew. 2010. "Presidential Leadership of the Tone of News Coverage and Public Opinion on Iraq" (paper presented at the annual meeting for the American Political Science Association, Washington, DC., August 25, 2010).

Judita, P., Mark S., and Robert G. 2015. "Exploring Relation Types for Literature Based Discovery." *Journal of the American Medical Informatics Association* Vol. 22, No. 5: pp. 987-992.

Krippendorff, Klaus. 2004. *Content analysis: an introduction to its methodology*. CA: Sage Publication.

Mills. C. Wright. 1956. *The Power Elite*. NY: Oxford University Press.

Ramos, Juan. 2003. "Using Tf-idf to Determine Word Relevance in Document Queries" (paper presented at the first instructional conference on machine learning) Vol. 242.

Rosenau, James N. 1971. *The Scientific Study of Foreign Policy*. NY: Free Press.

_____. 1976. *Comparing Foreign Policy: Theories, Findings and Methods*. NY: Sage Publications. p. 29.

Salton. Gerard, and Christopher Buckley. 1988. "Term-weighting approaches in automatic text retrieval" *Information Processing & Management* Vol. 24, No. 5.

Stigler, George J. 1971 "The Theory of Economic Regulation." *The Bell Journal of Economics and Management Science* Vol. 2, no. 1: pp. 3-21.

Strömbäck, Jesper, and Spiro Kiousis. 2010. "A new look at agenda-setting effects—Comparing the predictive power of overall political news consumption and specific news media consumption across different media channels and media types," *Journal of Communication*. Vol. 60, No. 2: pp. 271-292.

Preiss, Judita, Mark Stevenson, and Robert Gaizauskas. 2015. "Exploring relation types for literature-based discovery," *Journal of the American Medical Informatics Association*. Vol. 22, No. 5: pp. 987-992.

Waltz, Kenneth N. 1959. *Man, the State, and War: A Theoretical Analysis*. NY: Columbia University Press.

Big Data Analysis of Foreign Policy with a Focus on the Speeches by Ministers of Foreign Affairs in Park Geun-hye and Moon Jae-in Administration

Sion Lee

This paper examined the changes of foreign policy during Park Geun-hye and Moon Jae-in government with official speeches of the Minister of Foreign Affairs in each administration. The results were as follows. First, some keywords were actively used in both administrations such as ‘North Korea’, ‘Korean peninsula’, and ‘peace.’ Meanwhile, the Park administration preferred the words including ‘Northeast Asia’, ‘economy’, ‘security’, and ‘reunification’ while the Moon government often used specific words such as ‘people’, ‘human rights’, ‘sustainability’, and ‘communication.’ Next, based on the results obtained by LDA (Latent Dirichlet Allocation) analysis, the diplomatic policies of the Park government could be summarized as 8 topics while the major foreign policy agendas of the Moon administration were identified as 6 topics, which indicates the difference in foreign policies between the two governments. The study also showed that the policy priorities might differ over time within a single government.

Keywords: Big Date Analysis, Public Discourse, Korean Foreign Policy, Diplomatic Strategy, Text Mining, Topic Model