



한미 정권교체에 따른 주한 미대사관 외교문서의 주제와 감정표현 변화

위키리크스 공개 외교전문의 컴퓨터 언어처리 분석*

백영민**

연세대학교 언론홍보영상학부 조교수 ymbaek@yonsei.ac.kr

최문호**

서울대학교 언론정보학과 박사과정 bird@kbs.co.kr

장지연***

한국과학기술원 전산학과 박사과정 ssiru87@gmail.com

이 연구는 위키리크스를 통해서 공개된 주한 미국대사관의 외교전문들을 컴퓨터 언어처리 기법을 이용하여 분석한 것이다. 구체적으로 2006년 1월 1일부터 2010년 2월 25일까지 미국대사관에서 본국으로 발송한 1,080건의 ‘비밀’ 혹은 ‘대외비’ 외교전문들을 분석하였다. 해당 시기는 본국인 미국과 주재국인 한국 모두에서 정권교체가 일어났던 시기였다. 한국의 정권교체를 ‘대등적-협력적’로 미국의 정권교체를 ‘일방적-상호적’로 단순화한 후 위키리크스에서 공개된 전문들을 ‘대등적-일방적’, ‘협력적-일방적’ 그리고 ‘협력적-상호적’의 세 기간으로 나누었다. 또한 공개된 외교전문들에서 ① 한국 내 정세, ② 북한관련, ③ 주한미

* 본 논문의 일부는 2012년 한국언론학회 정치커뮤니케이션분과 학술회의에서 발표되었습니다. 당시 본 논문을 읽고 유익한 조언을 해 주신 성균관대학교 정성은 교수님, 서울대학교 이준용 교수님과 한규섭 교수님 그리고 언론정보연구의 익명의 심사위원님 세 분께 이 자리를 빌어 진심으로 감사드립니다.

** 교신저자

균, ④ 남북경협, ⑤ 해외파병, ⑥ 동북아문제, ⑦ 한미통상의 주제들을, 그리고 긍정적, 부정적 감정과 두 감정이 공존하는 양가감정(兩價感情)을 각각 잠재적 디리실레분포 배치모형, 주성분 분석, 그리고 LIWC 분석을 통하여 추출하였다. 분석결과, 첫째, 주제와 감정을 추출하는 컴퓨터 텍스트 분석은 일정수준의 타당성을 확보할 수 있었고, 둘째, 문서내의 주제와 감정의 변화는 주로 본국의 정권 변화에 민감하게 반응하며, 주재국의 변화에는 그 변화가 상대적으로 적다는 사실을 확인할 수 있었다. 이 논문에서는 위키리크스 외교전문과 같은 대용량의 텍스트를 분석할 때 컴퓨터 언어처리 방식의 방법론적 타당성과 주재국-대사관-본국간의 외교 커뮤니케이션의 이론적 함의도 논의하였다.

KEYWORDS 위키리크스 • 컴퓨터 텍스트 분석 • 잠재적 디리실레분포 할당모형 • 주성분분석 • 감정분석 • 시계열분석 • 외교 커뮤니케이션

“외교적 목적을 위한 본국과의 공식서신들은 절대 침해받지 않는다.”

- 외교관계에 관한 비엔나 협약, 제27조 2항

(The official correspondence of the mission shall be inviolable.

- Vienna Convention on Diplomatic Relations, 1961, Article 27. 2)

1. 연구의 목적

대사관은 주재국과 본국과의 외교적 커뮤니케이션의 창구역할을 수행한다. 외교적 목적을 위한 본국과의 공식적 서신들은 국제적 관례를 통해서 또한 협약을 통해서 철저히 보장되고 있다. 따라서 대사관의 비밀 서신들은 본국의 대리인(delegate)이 본국의 이익과 관련한 주재국 평가를 보여주고 있으며, 따라서 주재국에 대한 대외정책을 수립 및 실시할 때 본국이 가장 중요하게 반영하는 자료가 된다. 미국의 경우 1966년부터 실시된 정보공개법(Freedom of Information Act)에 따라 비밀보호기간

이 해제된 외교문서들의 경우 열람이 가능하였다. 따라서 1차 자료, 즉 정부공식문서에 기반한 외교문제 혹은 국제문제를 연구하는 학술적 연구 대부분은 보통 30년 이상의 사건들에 대한 문헌분석을 시도하고 있다 (예를 들어, 홍석률, 2004; Cummings, 1981/2001 등 참조).

그러나 2010년 11월 28일 위키리크스가 약 25만 건의 외교전문(外交電文)을 인터넷으로 공개하면서 상황은 뒤바뀌었다. 오바마 행정부의 법률 고문인 고흥주 변호사의 말(Koh, 2010. 11. 27)처럼 위키리크스의 외교전문의 공개는 당시진행 중이었던 미국의 대외전쟁, 그리고 미국과 동맹국들의 관계를 위기에 처하게 만들 수 있는 부분이 적지 않았다(또한 영국주재 미국대사인 Susman, 2010. 11. 28의 논평도 참조하라). 실제로 위키리크스의 외교전문들이 공개된 후, 튀니지의 부패한 대통령인 벤 알리(Ben Ali)는 권좌에서 축출되었는데, 위키리크스에서 공개된 정보들은 중동과 북아메리카에 일었던 재스민 혁명(Jasmine Revolution)에 큰 영향을 미친 것으로 알려져 있다(김유향·조희정, 2011; Kim, 2011). 실제로 국내에서도 이명박 전 대통령과 관련하여 친형인 이상득 의원의 “이명박 대통령은 철두철미하게 친미(親美)·친일(親日)(President Lee was pro-U.S. and pro-Japan to the core)”이라는 언급은 미국산 소고기 수입이나 한·미 자유무역협정(Free Trade Agreement)과 같은 국내적 이슈와 맞물려 언론 및 인터넷 공간에서 큰 파장을 불러오기도 하였다.

그러나 위키리크스의 외교전문의 파급력에도 불구하고, 경험적 분석연구는 의외로 적다. 위키리크스에 관한 해외 및 국내 연구는 대개 메타적 수준에서 위키리크스를 분석하고 평가한 문헌이 대부분이다. 위키리크스에 관한 문헌들은 크게 세 갈래로 분류할 수 있다. 첫 번째는 위키리크스의 등장에 따른 표현의 자유와 같은 가치와 국가안보와 같은 가치를 어떻게 저울질하는 것이 옳은가에 관한 규범적(normative) 연구다 (이진우, 2011; 이태훈, 2010; 조동원 2011; Fenster, 2011; Heisbourg, 2004; Oper, 2011; Page & Spence, 2011). 이런 연구는 위키리크스뿐 아니라 다른 정보공개 사이트들(이를 테면 OpenLeaks, Cryptome 등)

전반을 다루고 있으며, 정보공개에 합리적 범위는 어느 정도여야 하는지, 어떤 상황에서 관련 정보의 공개를 금지하여야 하는지, 어떤 법률적 근거로 국가안보에 위협이 되는 공개행위를 처벌해야 하는지 혹은 처벌행위를 하는 것이 민주주의적 가치와 양립할 수 있는 것인지 등을 다루고 있다. 이런 연구의 출발점은 위키리크스의 현실적 파급력을 강조하는데, 대부분의 논의들이 외교전문 전반에 대한 경험적 분석보다 논란이 된 일부 표현과 전문을 중심으로 지나치게 규범적인 성격을 보이고 있다는 문제가 있다. 그러나 라크만(Rachman, 2010, 12, 13)의 언급처럼 위키리크스의 내용들 대부분이 미국 정부가 공식적으로 밝혔던 사실들과 부합하며 외교관계에 미쳤던 실제적인 효과는 상대적으로 아주 미미하다는 점, 그리고 위키리크스로 인해 붕괴한 정권들은 모두 국민들의 표현의 자유를 억압해 왔던 권위주의 국가들이며 민주주의에 기반한 정부들은 별다른 영향을 받지 않았다는 점(Kim, 2011)을 감안할 때, 정보공개에 범위에 대한 논의는 “XX에 대한 표현은 처벌받아야 한다”는 결론으로 이어질 가능성이 크다. 즉 위키리크스가 폭로했던 내용에 대한 경험적 분석 없는 규범적인 잣대의 적용은 자칫 국가 안보를 위한 논의로 변질되면서, ‘표현의 자유’라는 민주적 가치를 훼손할 위험성도 부정할 수 없다.

두 번째 갈래는 위키리크스가 국내외 정치세력들 사이의 권력역학에 미친 영향에 관한 연구이며, 주요 주제는 위키리크스가 중동 및 북아프리카 정세에 미친 영향이다(김유향·조희정, 2011; Fenster, 2011; Goldstone, 2011; Kim, 2011). 튀니지의 재스민 혁명 사례가 위키리크스의 정보가 가져온 변화의 사례로 가장 많이 언급되고 있다. 이 외에도 중동의 대표적인 친미정권인 사우디 아라비아의 압둘라(Abdullah) 국왕이 이란의 핵 개발 방지를 위해 미군의 군사적 개입을 요청했다는 위키리크스의 공개를 빌미로, 이란의 마무드 아마디네자드(Mahmoud Ahmadinejad) 대통령은 정치적 경쟁자인 마노체르 모타키(Manouchehr Mottaki) 외무장관을 제거하기도 하였다(보다 풍부한 사례로는 Kim, 2011, 23~25쪽을 보라). 그러나 이러한 연구는 과연 위키리크스의 정보가 그러한 변화를 직접적으로 일으킨 직접적 원인인가에 관한 질문에 대

한 과학적인 답을 주지 못하고 있다. 비근한 예로 우리나라의 경우도 독도와 관련한 미국대사관의 발언이나, 이명박 전 대통령 및 한-미 FTA 등과 관련된 언급들이 국내정치에 파괴력 있는 영향력을 끼치지 못했다. 결국 위키리크스의 효과에 관한 근거로 제시되는 사례들의 대부분은 관찰자들의 인상비평 수준에서 크게 벗어나지 못하며, 어떤 시각을 택하는가에 따라 다른 설명들이 충분히 가능하다는 문제점을 안고 있다.

마지막에 해당되는 연구들은 저널리즘의 관점에서 위키리크스 폭로행위를 살펴본 연구들이다(김헌태, 2011; 조동원, 2011; Benkler, 2011; Kim, 2011; Opper, 2011). 언론사의 가장 중요한 기능은 바로 공중(公衆)에게 정보를 전달하는 것인데, 위키리크스 또한 '폭로'를 통해 정보를 제공한다는 점에서 언론의 한 형태로 볼 수 있다. 특히 탐사보도 저널리즘 전통에서 어산지의 폭로는 영웅적 언론인과의 유사한 면이 없지 않다. 그러나 오퍼(Opper, 2011)는 위키리크스와 기존의 제도권 언론은 명백하게 다르다고 주장한다. 제도권의 저널리스트들이 자신의 폭로에 대해 책임을 지려는(accountable) 모습을 보이는 반면, 위키리크스를 통한 폭로는 사회적 책임을 지려 하지 않는다. 즉 위키리크스는 사회적 책임성(social responsibility)을 강조하는 제도권의 언론 모형(Siebert, Peterson, & Schramm, 1956)과 배치되기에 언론이라고 보기 어렵다. 비슷한 맥락에서 요카이 벤클러(Yochai Benkler, 2011)는 전문가주의(professionalism)와 책임(responsibility)이라는 기존 미디어의 가치와 거리가 있다는 점을 들어(p. 315) '위키리크스 폭로'와 '저널리스트의 폭로'를 구분하고 있다. 언론학이라는 관점에서 위키리크스의 성격을 밝히는 연구는 필요불가결하지만, 이 역시 저널리즘이라는 규범적 가치관을 정립하려는 시도라는 점에서 경험적 연구와 구분되어야 한다.¹

1 본 논문의 목적은 위키리크스 본문에 대한 경험적 분석이다. 저널리즘과 위키리크스를 연결 짓는 작업은 언론학에서 수행될 충분한 가치가 있다. 즉 규범적 판단(즉 위키리크스의 폭로를 저널리즘 행위로 볼 것인가의 문제)과 경험적 분석 두 가지를 연결 짓는 작업이 필수적일 것이다. 연구의 필요성에도 불구하고, 기존의 경험적 분석연구가 충분하지 않은 상황에서 두 가지

이런 점에서 이번 연구는 위키리크스에서 공개된 외교전문에 대한 최초의 경험적이고 포괄적인 연구라는 의의를 지닌다. 물론 위키리크스 문서들은 전산학 분야에서 경험적으로 연구되고 있기는 하지만 (Chandracsekaran, 2012; Mayer & Rauber, 2011), 그 전문에 담겨 있는 내용에 대한 사회과학적 탐구와는 거리가 멀다. 이들 연구들은 자신들이 개발한 알고리즘이나 자연어처리 기법의 효과성(effectiveness)과 효율성(eficiency)을 테스트하기 위한 텍스트로 이용하기 위해 위키리크스를 분석할 뿐, 분석된 자료를 통해서 사회적 의미를 뽑아내는 것이 목적이 아니라는 점에서 이번 연구와는 상당히 다르다.

이 연구에서 우리는 대한민국주재 미국대사관에서 본국으로 발송한 외교전문 중 비밀(secret)과 대외비(confidential)로 분류된 총 1,080건의 문건들을 분석하여, 해당문서들에서 어떠한 주제들이 나타났는지 그리고 문건의 작성자가 어떠한 감정적 평가를 내리고 있는지를 외교전문이 발송된 시점과 한국 및 미국의 주요 정치적 사건과 연동하여 분석하였다. 흥미롭게도 여기서 분석된 위키리크스 외교전문들은 한국과 미국의 정권 교체기와 맞물려 있다. 즉 한국의 경우 노무현 전 대통령의 참여정부에서 이명박 전 대통령으로 정권이 교체되면서, 정부의 이데올로기적 지형은 '진보(미국과 대등한 관계를 주장)'에서 '보수(미국과의 협력을 강조)'로 변화하였다.² 미국의 경우 전 대통령인 조지 부시(George W. Bush)의 공화당 정권에서 현 대통령인 버락 오바마(Barack H. Obama)의 민주당 정권으로 이데올로기적 지형은 '보수(미국의 가치와

를 연결시키는 작업을 한편의 논문으로 완성하는 것은 지난(至難)한 일이다. 이에 대해서는 향후 연구과제로 결론부에서 정리하였다.

2 여기서 말하는 진보-보수의 이데올로기적 지형은 해당 국가에서 일반적으로 이해되고 있는 이데올로기적 성향을 의미하며, 특정한 정치철학이나 이념에 근거하지 않은 것임을 밝힌다. 실제로 진보-보수의 구분은 상대적인 것이며, 참여정부의 경우도 소위 '신자유주의'라고 비판하는 통합진보당 입장에서 보수적이라고 볼 수 있다. 그러나 사회전반의 이데올로기적 지형에서 판단할 때, 참여정부가 현재의 정부에 비해서는 상대적으로 소위 '진보'에 가까우며, 오바마 행정부의 경우는 부시 행정부와 비교할 때 상대적으로 '진보'에 가깝다.

이익을 우선시하는 일방주의)’에서 ‘진보(국제공조와 협력을 상대적으로 강조하는 상호주의)’로 바뀌었다. 전체 1,080건에 달하는 대외비 및 비밀로 분류된 외교전문들은 최근 개발된 컴퓨터 언어처리 기법을 이용하여 분석하였다. 이러한 컴퓨터 언어처리 기법의 활용과 함께 외교전문이 작성된 시기에 대한 이론적 고찰을 병행 실시하여 정량적 텍스트 분석의 타당성을 확보하였다.

2. 외교관계를 둘러싼 이론적 논의

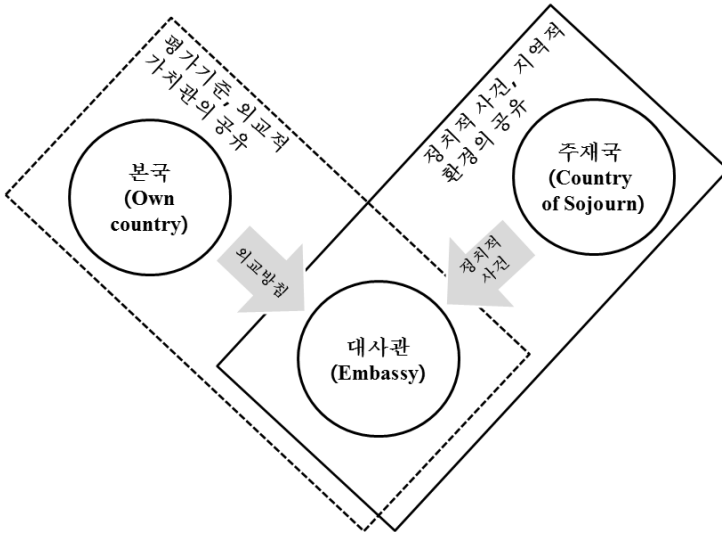
1) 외교 커뮤니케이션 맥락에서 대사관의 역할

대사관은 치외법권의 구역이며, 대사관과 본국의 연락은 주재국의 규제를 받지 않는다. 대사관은 본질적으로 본국의 이익을 위해서 존재한다. 따라서 외교사안을 접할 때 대사관은 본국과 관점 및 시각을 공유하게 마련이다. 그러나 대사관은 본국이 아닌 주재국과 공간적, 그리고 시간적 경험을 공유한다. 외교 커뮤니케이션 맥락에서 대사관이 처한 상황은 <그림 1>과 같다.

주재국을 바라보는 시선은 본국의 외교방침이나 본국의 국익을 따른다. 따라서 일어난 사태를 어떻게 바라보며 어떻게 해석할 것인가에 대한 가치판단은 본국의 외교방침의 변화에 따라 변화할 것이다. 그러나 자신들이 보고해야 하는 주재국의 정치적 변동사항들은 주재국의 주권에 달려 있으므로 오직 관찰의 대상이 될 뿐이다. 따라서 대사관에서 본국으로 보내는 외교전문들은 “본국의 시선으로 본 주재국의 상황변화”라고 정의할 수 있다.³

3 본문에서 제기하는 외교 커뮤니케이션은 언론학적 관점에서 살펴본 주재국의 사건, 외교문서 작성, 그리고 본국에 의한 외교문서 독해 과정을 의미한다. 따라서 외교 커뮤니케이션은 ‘외교’관계를 연구하는 외교학이나 국제관계학, 역사학 등의 사회과학분과와 경쟁관계에 있지 않

그림 1. 대사관의 위치와 외교커뮤니케이션



2) 미국의 정권교체가 미대사관 커뮤니케이션에 미치는 영향

모든 나라에서 정권교체는 국정운영 방향 및 기조 변화의 신호탄이다. 양당제가 정립된 미국의 경우, 실질적으로 유권자는 공화당과 민주당 중의 하나를 선택할 수밖에 없고, 소수당에 대한 투표는 사표(死票)에 가까운 것이 현실이다. 이러한 미국의 양당제 정당구조는 듀버거의 법

다. 대신 본 논문에서는 외교 커뮤니케이션을 내용분석의 대상으로 파악한 후 다른 사회과학 분과의 관점과 다른 언론학적 관점을 동원할 때 얻을 수 있는 장점이 무엇인지를 밝히는 것에 초점을 맞추었다. 본 논문의 관점은 ① 포괄성과 ② 상대적 중립성이라는 점에서 강점이 있다. 본 논문은 특정한 이론적 관점이나 개념들에 얽매이지 않고 외교문서의 전반적 흐름을 특정한 이론적 입장을 취하지 않은 채로 연구자가 중요하다고 판단한 개별 문서가 아닌 문서전체를 파악할 수 있다. 이를테면 국제관계론에서는 현실주의(현실주의 내에서도 여러 갈래가 있으나 지면의 한계로 여기서는 다루지 않는다), 자유주의, 구성주의 등을 학자들이 취하는 입장에 따라 취합되고 강조되는 자료가 달라지기 쉽다. 그러나 본 논문처럼 전수(全數)의 외교문서를 살펴볼 경우 이론적 관점에 의한 선별적 자료선택이라는 문제는 상대적으로 약하다.

칙(Duverger's Law) 혹은 합리적 선택이론(Rational choice theory)으로(Downs, 1957; Riker, 1982) 혹은 미국의 정치적 안정성을 도모한 파워엘리트들의 담합에 의한 것으로(Gitlin, 1980) 설명되기도 한다.

공화당과 민주당의 정책노선이나 방향은 상당히 다르다. 미국 내의 정권교체가 미국 내의 정치 및 경제에 미치는 영향을 시계열적으로 분석한 바르텔(Bartel, 2008)은 미국 내의 정치적 경제적 불평등의 수준이 공화당 집권기에서는 심화되지만, 민주당 집권기에서는 완화되고 있다는 증거를 제시하였다. 유사한 맥락에서 길리건(Gilligan, 2011)은 공화당에 비해 민주당 정권에서 보건관련 지표(자살건수와 살인건수)가 뚜렷하게 개선된다고 주장한다. 즉 정권교체는 정부 정책방향에 영향을 미친다. 그렇다면 정권교체는 외교정책에 어떠한 변화를 가져오는가?

두 정당의 이데올로기적 차이가 외교정책의 차이를 만들어 낸다는 학자들은 다음과 같은 몇 가지 근거를 제시한다. 첫째, 미국의 국내정책과 대외정책은 서로 분리되어 있는 것이 아니라 상호연관되어 있다. 이를테면 팽창적 대외정책을 추구하기 위한 비용은 결국 국내정책의 일부를 희생하며 얻게 마련이다. 대표적인 사례로 냉전시대 로널드 레이건 대통령의 군비확장정책, 그리고 부시(George H. W. Bush) 대통령의 걸프 전쟁을 들 수 있다. 공화당출신 대통령들의 이러한 정책은 미국 내 저소득층을 위한 복지정책에 대한 희생 혹은 재정적자를 감수하면서 이루어진 것이다(Ambrose, 1993/1996). 실제로 아이젠하워(Dwight Eisenhower) 대통령은 군장성 출신임에도 불구하고 “우리가 만드는 모든 총, 진수시키는 모든 전함, 발사하는 모든 로켓은 결국은 굶주리되 먹을 것을 공급받지 못하는, 헐벗되 입을 것을 공급받지 못하는 사람들로부터의 도둑질을 의미한다”(Ambrose, 1993/1997, 164~165쪽 재인용)라는 말을 남긴 적이 있다. 다시 말해 소위 ‘일방주의’적인 외교정책을 실현하기 위해서는 재원이 필요하며, 결국 미국민을 위한 정책자원을 소모시킬 가능성이 높아진다. 따라서 사회복지(social welfare) 등을 강조하는 민주당 정권은 대외정책에 소요되는 비용을 감소시켜 국내정책을 위한 재원을 마련하는 것에 중점을 두는 대안적 대외정책을 추구하려는 유인을 가

지며, 이는 상대국가와의 외교적 협력을 강조하며 군사적 갈등을 방지 하는데 중점을 두게 된다. 즉 민주당은 ‘상호주의’적인 외교를 구사하려 시도한다. 다시 말해 두 정당이 각각 추구하는 국내정책에 따라 이들의 대외정책은 ‘공화당-일방주의’ 그리고 ‘민주당-상호주의’로 단순화시켜 설명할 수 있다.

둘째, 대외정책을 주관하며 대통령에게 정책적 조언을 제공하는 참모들이 갖고 있는 보수적 혹은 진보적 가치관이 정책결정에 중요한 영향을 미치기 때문이다. 비록 대통령이 최종적 결정권자이지만, 전문지식으로 무장한 참모의 관점이나 조언은 결코 무시할 수 없다. 대표적 사례로는 부시(George W. Bush) 대통령에 대해 체니(Dick Cheney), 럼스펠드(Donald Rumsfeld) 등 소위 네오콘(neo-conservative) 참모들이 대외정책 결정에 미친 영향을 들 수 있다. 물론 9·11 테러리스트 공격과 같은 사건들을 무시할 수는 없겠지만, 네오콘들에 의해서 추진된 정책들은 미국의 압도적 군사력에 근거한 일방주의적 외교로 나타났다(윤영관, 2009). 실제로 대통령의 참모진이 대외정책의 방향성을 결정짓는 사례는 네오콘 등장 이전의 미국 외교사에서 종종 나타나는 현상이다(George, 1980).

셋째, 선출된 권력은 지지기반이 되는 국민의 소리를 무시할 수 없기 때문이다. 대표적으로 언급되는 사례가 카터(Jimmy Carter) 대통령이다. 워터게이트(Watergate)라는 도덕적 추문의 여파로 1976년의 대통령 선거의 주요 관심사는 정치인의 도덕성(morality)이었다. 실제로 카터 대통령은 자신이 도덕적임을 항상 강조하였고, 대외정책의 핵심어로 인권(human rights)을 채택하였다(McElroy, 1992). 일례로 카터 행정부는 당시 한국의 박정희 정권과 인권이라는 주제를 놓고 대립하기도 하였다(마상운·박원근, 2009). 또한 공화당과 민주당의 지지세력들 역시 대외정책에 대한 상이한 시각을 보여주고 있다(Zaller, 1992). 즉 핵심적 지지세력들은 선출된 지도자가 자신들의 가치관과 배치되는 정책을 실시할 경우 강하게 반발하여 행정부의 정책집행 능력을 감소시킨다. 실제로 아프가니스탄 전쟁을 계속 수행하고 있는 오바마 행

정부에 대해 2008년 선거 당시 그의 열성적 지지자였던 영화감독 마이클 무어(Michael Moore)는 오바마 정부 등장 후 대외정책과 해외파병에 변화가 없었다는 점에 대해 극도의 강한 실망감을 표시하기도 하였다(www.michaelmoore.com).

물론 미국 행정부의 이데올로기적 차이가 대외정책의 이데올로기적 차이로 언제나 연결되는 것은 아니다. 이를테면 보수적인 공화당 출신의 닉슨(Richard Nixon) 대통령은 공산주의 국가인 중공과 수교를 시작하기도 하였으며, ‘군산복합체’라는 말로 유명한 군장성 출신의 아이젠하워 대통령은 군비증강에 대한 비판적 태도를 견지하고 베트남의 공산화에 대해 강경대응을 피했다. 반면 진보적 민주당 출신의 케네디(John F. Kennedy) 대통령은 공산화를 저지한다는 명분으로 프랑스를 대신하여 베트남전을 개시하는 등 일방주의적 모습을 견지하기도 하였다. 베트남 전쟁에 대한 미국 미디어의 보도와 시민사회의 동향을 세밀하게 분석한 할린(Hallin, 1986)은 미국의 정권교체와 베트남전 정책이 별 상관이 없었다는 점을 강하게 강조하고 있다. 할린에 따르면 베트남전에 관한 정책에서 보수-진보의 의견대립이 나타나기 시작한 것은 정책결정자들의 주관적 관점이 아닌 전사자 증가와 월맹군의 구정(舊正) 공세를 전후로 벌어진 전장(戰場) 상황이었다고 한다. 미디어를 통해서 드러난 반전 움직임 역시도 베트남 전쟁을 둘러싸고 통치 엘리트들 사이의 균열이 발생한 이후에야 나타났다(Bennett, 1990; Hallin, 1986).

국내적 정치지형에서의 보수-진보 이데올로기로 완벽하게 환원되기는 어려운 여러 사례들에도 불구하고, 미국 행정부의 대외정책이 정권교체에 따라 대략적으로는 변화한다는 것은 어느 정도 인정되는 사실이라고 할 수 있다(기타 외국의 사례로는 McCraw, 1994 참조). 만약 그렇다면 주재국내의 대사관에서 주재국을 바라보는 시선 또한 본국의 정책변화 기조에 맞게 변화할 가능성이 높다고 볼 수 있다. 따라서 우리가 연구대상으로 설정한 위키리크스의 경우, 부시 전 대통령에서 오바마 현 대통령으로 정권이 바뀔에 따라 본국으로 보낸 전문들의 주제나 감정 표현이 달라질 가능성을 고려해 볼 수 있다.

3) 주재국의 정권교체가 대사관의 커뮤니케이션에 미치는 영향

앞서의 설명이 미국의 정권교체를 중심으로 외교노선의 변화를 설명하였다. 이번에는 주재국의 정권교체에 따라 미국의 외교노선이 어떻게 변화하는가를 살펴보기로 한다. 1차·2차 세계대전을 거치면서 미국은 고립주의 외교정책에서 탈피하고 세계의 열강으로 등장하였으며, 미국은 자신이 주도할 수 있는 세계질서 재편의 관점에서 타국의 정권교체를 바라보고 있다(Ambrose, 1993/1997). 자국의 이익을 주재국에 관철시키는 것은 많은 비용과 노력이 소모되는 일이며, 또한 국제적 신망을 떨어뜨릴 수 있기 때문에 가능하면 피하는 것이 합리적이지는 않지만(Nye, 2006), 2차 세계대전 이후에 미국은 심지어 자신들의 외교정책에 걸림돌이 되는 외국정권을 공작과 쿠데타 지원과 같은 비합법적 방식으로라도 전복시켰다. 흔히 급진주의 이론가들이 지적하는 신제국주의가 바로 이것이며, 이러한 예로 남베트남의 부패정권이자 베트남전을 수행하는 데 걸림돌이 되었던 고 딘 디엠(Ngo Din Diem) 정권의 전복(Ambrose, 1993/1997, p. 244), 니카라과의 좌익 정권의 붕괴를 시도했던 콘트라(contra, contra-revolution) 활동(같은 책, p. 371) 등을 들 수 있다. 실제로 뉴딜정책으로 미국의 사회보장제도의 초석을 닦았을 정도로 진보적이라는 평가를 받고 있는 루즈벨트(Franklin D. Roosevelt) 대통령도 부패로 악명이 높았던 니카라과의 소모자(Anastasio Somoza)에 대해서 “소모자는 개자식이지만, 그래도 우리 편인 개자식이다(Somoza may be a son of a bitch, but he’s our son of a bitch)”라는 말을 남기기도 하였다. 이 같은 발언은 미국의 입장에서는 언제나 ‘친미적’ 정권을 선호한다는 것을 방증한다.

따라서 미대사관의 입장에서는 주재국이 미국에 적극적으로 협력하려는 노선을 표방하는 정권이 등장하면 긍정적인 태도를 취하지만, 그렇지 않은 정권이 등장하면 부정적인 태도를 취할 것으로 예측할 수 있다. 우리가 연구대상으로 설정한 위키리크스 외교전문의 경우, 노무현 전 대통령에서 이명박 전 대통령의 정권교체 기간을 포함하고 있다.

노무현 전 대통령의 경우 이라크 파병에서 보이듯 ‘반미’까지는 아니라고 할 수 있어도, “반미면 어쩐가?”라는 말이 상징하듯 이명박 전 대통령에 비해 대미외교를 중시하는 성향이 그다지 크지 않다고 보는 것이 타당하다. 즉 한국에서 미국과의 협력을 매우 중시하는 성향을 보이는 정권이 등장하기 전·후에 미대사관의 외교전문의 토픽이나 논조가 바뀔 수 있다.

4) 본국-주재국의 정치변동이 외교 커뮤니케이션에 미치는 영향

앞서의 논의들은 미국, 혹은 주재국의 정권변화만을 일방적으로 고려하고 있다.⁴ 그러나 정권변화 외에도 외교 커뮤니케이션에 영향을 미치는 여러 요인들을 생각할 수 있다. 미시적 요인으로는 국가수반의 개인적 성향(Overy, 2004)을, 맥락적 요인으로는 본국-주재국을 둘러싼 주변 국가의 정세를 고려할 수 있다. 하지만 전통적으로 국제관계학이나 외교학에서는 본국-주재국 사이의 권력관계를 가장 중요한 요인으로 꼽는다(Keohane & Nye, 1977; Morgenthau, 1948/1987; Nye, 2006; Waltz, 2000). 비록 “힘의 불균형(asymmetric power)”이 국제관계의 가장 큰 특징이라고 하지만, 당사자들 사이의 상호의존을 통해서 보다 효과적인 권력운용이 가능하다(Keohane & Nye, 1977). 특히 달러약세, 막대한 전비와 재정적자 등의 문제를 안고 출범했던 오바마 행정부는 미국 일방주의적 외교정책을 실시하는 것이 매우 어려운 상황이었다. 따라서 오바마 대통령의 경우 이념이나 군사적 압박을 중요시하는 하드파워(hard power)보다는 다자간의 협상과 대화에 보다 중점을 두는 소프트파워(soft power)를 중시하는 모습을 보였다(윤영관, 2009). 다시 말해서 미국의 경우 자국의 이익과 이념을 앞세우는 “일방주의”정책기조에서, 상대적으로 주재국과의 대화와 협력을 통한 조화를 중시하는 “상호주의”

4 아쉽게도 본 연구에서는 정권변화를 제외한 다른 요인들을 고려하지 못하였다. 후속연구에서는 이러한 요인을 적극 고려하는 것도 흥미로운 발견으로 이어질 수 있을 것이다.

정책기조로 변화하고 있다고 단순화시킬 수 있다.

반면 미국의 대외정책기조의 변화와 맞물린, 한국의 경우는 상황이 미묘하게 엇갈리고 있다. 남북간의 대치상황과 최근 급성장하고 있는 중국에 대한 견제, 그리고 주한미군의 존재를 고려할 때, 미국이 바라보는 한국은 일반적 서유럽의 동맹국과는 다른 '관리의 대상'이 되는 동맹국일 수밖에 없다. 이러한 미묘한 입장에서 '대등한 관계'를 강조하는 정권에서 '미국과의 협력을 강조하는' 정권으로의 이동은 미국이 어떠한 정책기조를 택하는가에 따라서 다른 평가를 받기 쉽다.

우선 미국이 일방주의적 정책기조를 갖고 있는 상황에서 대등한 관계를 표방하는 한국정부는 불편한 대상이 될 수밖에 없다. 그 이유는 미국의 일방적 대외정책 추진에 주재국, 즉 한국정부가 걸림돌로 작용할 여지가 크기 때문이다. 그러나 미국의 입장에서 대미외교를 중시하며 미국과의 협력을 강조하는 한국정부는 매우 환영받을 대상이다. 그 이유는 소위 '불량국가(rogue state)' 중의 하나인 북한에 대한 미국의 대외정책 추진에 한국정부가 협조할 가능성이 매우 높아지기 때문이다.

반면 미국이 상호주의에 입각한 정책기조를 강하게 띠는 상황에서 대미외교를 중시하는 한국 정부가 편할 수도 혹은 불편할 수도 있다. 우선 미국과의 협력을 강조하는 한국정부는 미국의 외교정책에 적극적으로 협력적 자세를 보인다는 점에서 편한 상대일 수 있다. 그러나 한국의 보수층들이 대미관계의 중요성을 강조하며 미국과의 협력에 방점을 찍는지 살펴보면 도리어 불편한 외교상대로 인식될 수 있다. 전통적으로 한국의 보수층들은 북한의 도발을 방지하고, 더 나아가 북한을 압박하기 위한 군사적 혈맹(血盟)으로써 미국의 역할을 기대하기 때문에 대미외교의 중요성을 강조하여 왔다. 여기서 중요한 사안은 북한과의 관계를 바라보는 미국의 시각이다. 만약 미국이 북한과의 대화를 추진하려 한다면, 대미협력을 강조하는 한국정권은 협력을 구하기 까다로운 파트너가 되는 반면, 북한에 대한 압박을 추구한다면 더할 나위가 없이 좋은 파트너가 될 것이다. 그러나 취임시 벌어지고 있었던 이라크 및 아프가니스탄에서의 전쟁, 그리고 누적된 재정적자로 인해 중국과의 선린관계를

고려해야 하는 오바마 행정부의 입장에서 북한과의 갈등은 그다지 바람직한 외교정책이라고 할 수 없었다(윤영관, 2009). 즉 미국은 이명박 정부를 부담이 되는 파트너로 인식할 가능성도 배제하기 어렵다.

이러한 관점에서 한국대사관발(發) 위키리크스 외교전문들은 한국과 미국의 정권교체를 전·후하여 3개의 시기를 구분하였다. 우선 첫 번째 시기는 한·미의 정권을 ‘대등적-일방적’으로 요약할 수 있는 시기이다. 구체적으로는 한국의 대선이 끝나고 이명박 후보자가 차기 대통령으로 당선이 확정되기 전 시점까지 본국으로 보내진 위키리크스 외교전문들을 여기에 분류할 수 있다. 두 번째 시기는 ‘협력적-일방적’으로 요약할 수 있는 시기다. 이명박 후보자의 당선 시점과 오바마 후보자가 당선이 확정되었던 시점 사이에 작성되었던 외교전문들이 두 번째 시기에 포함된다. 세 번째 시기는 ‘협력적-상호적’으로 요약할 수 있는 시기로 오바마 후보자의 대통령 당선 확정 후에 공개된 위키리크스 외교전문들을 여기에 분류할 수 있다. 위의 논의 전개과정에 근거하여 우리는 다음과 같은 두 가지의 연구문제를 설정하였다.

- 연구문제 1: 세 시기에 걸쳐 외교전문에 실린 주제들은 어떻게 변화하고 있는가?
- 연구문제 2: 세 시기에 걸쳐 외교전문에 실린 감정의 표출은 어떻게 변화하고 있는가?

3. 연구방법

1) 분석대상 외교전문 선정

위키리크스가 공개한 외교전문들 중 1,980개가 대한민국 미국대사관에서 본국으로 보내졌다. 공개된 외교전문들은 1996년 12월 28일부터 2010년 2월 28일간 대사관과 본국과의 연락이었다. 그러나 대상이 되는

1,980개의 전문 중에는 노태우, 김영삼, 김대중 전 대통령 시절에 보내졌던 전문들도 있다. 그러나 노무현 전 대통령 취임 이전의 문서들은 그 양으로 볼 때 미미하며(총 9건), 연구문제의 점검 및 가설검증에 맞지 않아 분석에서 제외하였다.

공개된 미국의 외교전문은 그 기밀 정도에 따라서 비밀(secret), 대외비(confidential), 비분류일반(unclassified) 문서의 3종류로 나뉜다. 남은 전문들 중에서 6퍼센트 정도가 ‘비밀’로, 49퍼센트는 ‘대외비’로 그리고 45퍼센트는 ‘비분류일반’으로 분류되었다. 이는 위키리크스를 통해서 공개된 총 25만 건의 보안비율과 조금 다르다. 하이스버그(Heisbourg, 2011, p. 209)에 따르면 공개된 전체 외교전문은 경우 비밀등급이 6퍼센트, 대외비가 40퍼센트, 그리고 나머지 54퍼센트가 비분류일반이었다고 한다. 즉 한국대사관에서 본국으로 보내진 전문은 ‘비밀’ 문서 비율은 상당히 유사하지만, 전체적으로는 대외비의 비율이 상대적으로 높음을 알 수 있다. 한국 대사관에서 발송한 외교전문에서 대외비가 상대적으로 높은 이유는 아마도 남북대치라는 한반도의 특수한 상황과 주한미군의 존재 때문일지도 모른다. 우리들은 우리와 비슷한 방식으로 위키리크스 외교전문에 접근한 외국의 사례를 찾지 못하여 여기에 대한 자세한 의견을 담기는 어렵다. 우리는 이중 비밀의 수준이 가장 낮은 비분류일반에 해당되는 문서들은 분석에서 제외하였는데, 그 이유는 해당 문서들의 내용이 이미 언론에 공개된 내용을 단순하게 정리한 것이기 때문이다.

전술한 세 시기별 비밀과 대외비에 해당되는 외교전문들의 빈도는 <표 1>과 같다. 전반적으로 제 1시기에 해당되는 외교전문이 가장 많은데, 그 이유는 한국의 대선전에 외교전문들이 집중적으로 발송되었기 때문이다.

이번 연구에서는 이 중 ‘비밀’과 ‘대외비’에 해당되는 1,080개의 외교전문들을 분석 대상으로 삼아 각각의 문서들에 나타난 주제(topic)와 감정(sentiment)을 분석하였다.

표 1. 시기별 비밀 및 대외비 외교전문

	제 1시기 (노무현-부시)		제 2시기 (이명박-부시)		제 3시기 (이명박-오바마)		총합
	비밀	대외비	비밀	대외비	비밀	대외비	
빈도 (비율)	57 (5%)	502 (46%)	29 (3%)	190 (18%)	34 (3%)	267 (25%)	1,080 (100%)

알림. 세 시기와 비밀과 대외비의 비율은 통계적으로 유의한 관계가 나타나지 않았음.

$$(\chi^2_{(2)} = 1.49, p = .48)$$

2) 컴퓨터 언어 처리분석을 이용한 외교전문 주제분석

텍스트에 대한 컴퓨터 언어처리는 크게 '주제(topic)'와 주제에 대한 '감정적 방향성(valence)'을 분석하는 것으로 크게 양분할 수 있다. 컴퓨터 언어분석에서는 '주제'를 파악하는 방법을 '주제발견(topic detection)'이라고 부르며, '감정적 방향성'을 파악하는 방법을 '감정분석(sentiment analysis)'라고 일컫는다(Pang & Lee, 2008). 컴퓨터를 이용한 언어처리 기법을 소개하면서 4-1)에서는 주제발견과 관련된 기법을, 그리고 4-2)에서는 감정분석과 관련된 방법을 중점적으로 다루었다.

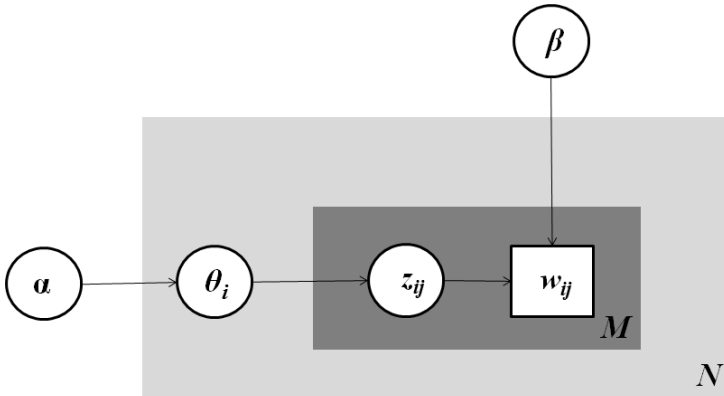
이번 연구에서는 주제발견의 한 기법으로 '잠재적 디리실레분포 할당모형(LDA, Latent Dirichlet Allocation)'을 통해 위키리크스의 개별 전문들에 내재되어 있는 주제들을 추출하였다. LDA는 생성적 확률 모형(generative probabilistic model)의 하나로 "말뭉치(corpus) - 문서(document) - 단어(word)"들의 조건부 확률(conditional probability)을 위계적 베이시안 모형(hierarchical Bayesian Model)을 구성함으로써 문서와 문서, 혹은 단어와 단어 사이의 유사성을 계산한 후 자료를 축약시키는 기법이다(Blei, Ng, & Jordan, 2003; Styvers & Griffiths, 2007). 여기서 생성적 확률모형이 의미하는 바는 전체 말뭉치에서 특정한 문서가 생성될 확률, 그리고 특정문서에서 특정단어가 쓰일 확률이 특정한 확률분포를 따른다는 것을 의미한다.

추상적인 설명 대신 예를 들어보자. 우리가 분석하려고 하는 문서들, 즉 주한 미국대사관에서 본국으로 보낸 전문들이 바로 ‘말뭉치’이고, 개별적 전문들은 ‘문서’가 되며, 개별 문서들은 ‘단어들’로 구성되어 있다. 즉 우리가 분석하려는 텍스트는 말뭉치-문서-단어의 순서를 갖는 위계적(hierarchical) 구조임을 알 수 있다. 이러한 위계적 구조는 최근 사회과학에서 인기를 끌고 있는 다층모형(multi-level model; 이에 대해서는 김두섭·강남준(2008)의 13장 참조)과 개념적으로 동일하다. 다층모형에서 흔히 제시되는 사례는 학교-교실-학생의 위계적 구조를 갖는데, 이는 텍스트에 내재된 위계적 구조와 동일한 형식을 가지며, 실제로 모수(parameter)를 추정하는 통계적 과정 역시 가정되는 확률분포가 다를 뿐 개념적으로는 동일하다(Blei, Ng, & Jordan, 2003, p. 997).

LDA는 ‘문서’ 수준과 ‘단어’ 수준에 하나의 잠재적 변수(latent variable)로서 ‘주제(topic)’를 상정한다. 다시 말해서 우리가 실제로 ‘관측 변수’인 ‘단어’들은 ‘잠재변수’인 ‘주제’를 구현하고 있다고 가정할 수 있다. 사실 주성분분석(principal component analysis), 군집분석(cluster analysis)과 같은 잘 알려진 사회과학 기법도 일반적으로 이러한 가정들에 기반하고 있다. 일례로 인자분석에서 관측변수인 ‘지표(indicators)’와 잠재변수인 ‘인자(factor)’ 사이의 관계는 LDA에서의 ‘단어’와 ‘주제’의 관계와 개념적으로 정확하게 동일하다(Styvers & Griffiths, 2007).

우리는 LDA에 필요한 모수들을 <그림 2>처럼 나타낼 수 있을 것이다. 아래의 설명부분은 LDA를 설명한 통계학 문헌(Blei, et al., 2003, pp. 996~997)에서 소개된 수식들이 의미하는 바를 사회과학통계분석의 관점에서 가능하면 평이하게 말로 풀어낸 것이다. 우선 사각형 M은 우리가 갖고 있는 위키리크스 외교전문들의 말뭉치를 나타낸다. 그러나 사각형 N은 위키리크스에 내재되어 있는 주제를 의미하며, 몇 개의 주제를 설정하는가에 따라 토픽의 범위가 달라진다. 주제들의 수가 적다면 하나의 주제는 넓은 의미의 단어들을 포괄하게 되며, 반대로 주제의 수가 많다면 해당 주제를 구성하는 단어들은 의미가 서로 유사하게 될 것이다. 인자분석이나 군집분석과 마찬가지로 주제의 숫자는 너무 많지

그림 2. 잠재적 디리실레분포기반 할당모형(LDA)에 대한 시각적 설명



알림. 사회과학문헌의 관례에 따라 관측되는 변수는 사각형으로, 잠재되어 있거나 혹은 추정되는 변수들은 원으로 표현하였다. 블레이 등(Blei, et al., 2003)의 연구에서는 w 역시 원으로 표기되어 있으나, 사회과학 문헌의 표기관례에 따라 원을 사각형으로 바꾸었음을 밝힌다. 모형 내부의 표기들과 설명들은 본문을 참조하라.

도, 또한 너무 적지도 않아야 한다. 너무 많을 경우는 간명한(parsimonious) 해석이 어려워지며, 너무 적을 경우는 지나치게 주제가 광범위하게 정의되어 해석의 엄밀성이 떨어진다.

N 이라는 공간에서 단어의 등장(측정변수)는 단어로 대변되는 주제(잠재변수)로 표현될 수 있다. 우선 전체 말뭉치에 I 개의 문서가 있고, 모든 문서들에 걸쳐 J 개의 단어가 있다고 가정하자. 또한 연구자가 K 개의 주제를 설정했다고 가정하자(여기서, 발견된 단어의 수보다 주제의 수가 적어야, 즉 $K < J$ 이어야 한다). 만약 우리가 단어를 w_{ij} 으로 그리고 주제를 z_{ij} 로 표기한다면, 바로 단어-주제의 관계는 N 이라는 사각형에 나타나있는 z_{ij} 와 w_{ij} 의 관계(즉 $z_{ij} \rightarrow w_{ij}$)로 나타낼 수 있다. 즉 w_{ij} 와 z_{ij} 는 모두 단어수준(word-level)의 변수다.

만약 하나의 문서를 하나 혹은 여러 개의 주제들을 포괄하고 있는 전체로 가정한다면 주제인 z_{ij} 는 개별 문서들의 특징에 따라 다르게 구현될 것이다. 여기서 개별 문서들의 특징을 θ_i 로 표현한다면 z_{ij} 의 발

현(發顯)은 θ_i 에 의해 조건지어지는, 다시 말해서 영향을 받는 모습(즉 $\theta_i \rightarrow z_{ij}$)으로 나타낼 수 있다. 즉 θ_i 는 그림처럼 바로 문서수준(document-level)의 공간인 M 에 위치하게 된다.

끝으로 우리가 분석의 대상으로 삼고 있는 말뭉치, 즉 위키리크스 문서들에 고유한 모수들을 고려할 수 있다. 언급하였듯 LDA는 말뭉치-문서-단어의 위계적 관계를 중심으로 전개되는 모형이기 때문에, 선택된 말뭉치의 고유한 모수들 역시 각 하위수준 별로 하나씩 존재하게 된다. 문서수준에서 적용되는 모수로는 α 를 그리고 단어수준에서 적용되는 모수를 β 라고 표시한다면, 각각의 변수들은 그림과 같이 $\alpha \rightarrow \theta_i$ 에, 그리고 $\beta \rightarrow w_{ij}$ 로 각각 나타낼 수 있다. 즉 α 와 β 는 말뭉치 수준의 변수다.

이렇게 볼 때, 만약 k 개의 주제가 있다고 가정할 때, 개별 단어가 k 개의 주제를 나타낼 확률, 즉 $\{(\beta, z_{ij}) \rightarrow w_{ij}\}$ 의 관계는 $p(w_{ij}|z_{ij}, \beta)$ 로 나타낼 수 있다. 또한 k 개의 주제를 나타내는 확률, 즉 $\{\theta_i \rightarrow z_{ij}\}$ 의 관계는 $p(z_{ij}|\theta_i)$ 로 표현할 수 있다. 따라서 N 의 공간(여기서는 위키리크스의 외교전문들 전체)에서 주어진 단어들을 이용해서 k 개의 주제를 파악하게 되는 확률들의 집합은 아래와 같이 표현할 수 있다.

$$\left(\prod_{j=1}^J \sum_{z_{ij}} p(z_{ij}|\theta_i) p(w_{ij}|z_{ij}, \beta) \right) \cdots \langle \text{공식 1} \rangle$$

여기서 말뭉치수준의 변수(α)가 주어졌을 때 개별문서들의 특정(θ)을 추정하는 확률공식은 $p(\theta_i|\alpha)$ 로 나타낼 수 있으며, 이를 위에서 언급한 <공식 1>과 통합한 후 개별 문서에 특정한 단어가 나타나게 될 확률을 도출하기 위해서 미분을 취하면 아래와 같은 <공식 2>를 도출할 수 있다.

$$p(w_{ij}|\alpha, \beta) = \int p(\theta_i|\alpha) \left(\prod_{j=1}^J \sum_{z_{ij}} p(z_{ij}|\theta_i) p(w_{ij}|z_{ij}, \beta) \right) d\theta \cdots \langle \text{공식 2} \rangle$$

위의 공식이 말하고 있는 것은 잠재변수 z_{ij} 을 가정한 말뭉치수준의 변수인 α 와 β , 그리고 개별 단어가 특정한 문서에서 나타나는 경험적 확률(w_{ij})만을 알면(즉 <공식 2>의 좌변을 보라), k개의 주제(즉 z_{ij})를 추정할 수 있다는 것이다. 개념적으로 인자분석과 동일한 데, 인자분석에서는 문서대신에 응답자를 그리고 단어들 대신에 개별 문항들을 투입한다는 것이 다를 뿐이다. 인자분석에서 적절한 인자의 수를 연구자가 자신의 판단에 근거하여 결정하듯이, 마찬가지로 LDA에서도 적절한 주제의 수를 연구자가 설정할 수 있다.

LDA에서 “디리실레(Dirichlet) 분포”라는 이름이 들어간 이유는 θ_i 를 뽑아내는 분포가 디리실레 분포를 따른다고 가정하기 때문이다. 다른 모수분포(이른테면 정규분포 혹은 이항분포 등과 같은)가 아닌 디리실레 분포를 가정하는 이유는 LDA의 분석이 베이지안 통계학에 기초하고 있기 때문이다. 베이지안 통계학에서는 사후분포(posterior distribution)를 사전분포(a priori distribution)와 데이터의 분포(data distribution)의 결합분포(joint distribution)로 가정하는데, 사후분포가 사전분포와 동일한 모수형태를 갖는 경우를 켈레 혹은 공액(共軛, conjugacy)이라고 부른다(Gelman, Carlin, Stern, & Rubin, 2004). 디리실레 분포를 가정할 때 좋은 점은 바로 디리실레 분포가 다항명목분포(multinomial distribution)⁵의 사전공액(conjugate prior) 분포를 가지기 때문이다(다시 말해 베이지안 통계학의 가정과 잘 어울리기 때문이다).

이번 연구에서 우리는 주제의 숫자를 총 50개로 가정하였다. 사실 전술하였듯 몇 개의 주제를 산출하는가는 탐색적 인자분석(EFA)과 마찬가지로 연구자의 자의적 판단이다. 물론 대안적으로 다른 개수의 주

5 사회과학에서 말하는 다항명목회귀분석은 결과변수가 범주형 자료일 때 사용하는데, 결과변수의 분포를 다항명목분포를 가정하기 때문에 붙여진 이름이다. 분야(이른테면 기계학습(machine learning)의 경우)에 따라서 다항명목분포 대신 범주형 분포(categorical distribution)라는 말을 사용하기도 한다.

제를 생각해 볼 수 있으나, 중요한 주제들의 숫자는 대략 10개 내외였으며, 아무리 폭넓게 잡아도 20개를 넘지는 않았다.⁶ 즉 50개 보다 더 많은 주제를 뽑는다고 하더라도 늘어나는 주제들은 언급된 중요한 주제들에서 크게 벗어나지 않는다. 이러한 이유로 우리는 한국대사관에서 본국(즉, 미국의 수도)으로 보내진 1,080개의 외교전문들이 총 50개의 주제로 구분될 수 있는 확률을 계산하였다. 이번 논문에서 50개의 주제들과 주제별 해당 단어들을 하나하나 다 설명하는 것은 현실적 어려움이 있다. 한국대사관에서 발송된 외교전문은 텍스트 포맷(*.txt)의 파일로만 총 12.8MB이며, 전체 단어의 수는 백만 개($N_{Total\ Words} = 1,081,856$) 그리고 고유 단어-전체 말뭉치에서 최소한 한 번 이상 나타난 단어-는 약 오만 개에 달한다($N_{Unique\ Words} = 49,012$).⁷

간략하게 인자분석에 비유해서 설명하자면, 데이터의 구조는 1080개의 사례수와 49,012개의 측정변수들이 있으며, 여기서 약 오만개의 변수들은 50개의 잠재변수들로 축약(dimension reduction)되었음을 의미한다. 정상분포(normal distribution)를 가정하는 인자분석과는 달리 LDA에서는 다항명목분포(multinomial distribution)를 가정한다. 다항 명목분포를 가정하는 이유를 예로 들면 beef라는 단어는 특정한 시기의 문서에서 높은 빈도를 보이는 반면(이를테면 이명박 전 대통령 집권초의 촛불집회), election이라는 단어는 다른 시기에 높은 빈도를 보일 것이다(이를테면 대통령 선거기간). 즉 개개의 단어는 모두 명목변수로 볼 수 있고, 이는 사회과학에서 결과변수가 명목변수일 경우 사용

6 실제로 주제의 개수를 다르게 설정한 후 주성분 분석(principal component analysis)을 실시하여도 선택되는 주성분의 개수는 아이겐값이 1.2 이상은 7개 내외, 그리고 통상적으로 이용되는 카이저 규칙(Kaiser rule), 즉 1.00 이상의 경우는 20개 미만으로 아주 일정하게 나오는 것을 발견할 수 있었다.

7 위키리크스 외교전문 및 LDA 분석결과를 참조하고 싶은 독자들은 저자에게 개별적으로 문의할 수 있다. LDA 분석결과와 경우 단어-주제(word-topic) 데이터와 문서-주제(document-topic) 데이터로 구분된다. 또한 50개가 아닌 다른 개수의 주제들을 선택한 경우도 저자를 통해 얻을 수 있다.

하는 다항명목 로짓(multinomial logit) 혹은 프로빗(probit) 회귀분석과 똑같은 가정을 공유하고 있다. 다항명목 로짓 회귀분석과 마찬가지로 하나의 사례(여기서는 문서)는 결과변수의 유무(여기서는 주제)를 포함하는 확률값을 갖게 된다. 즉 50개의 주제들이 얼마나 문서에 구현되었는가는 모두 50개의 확률값으로 나타나며, 따라서 문서별로 50개 주제들의 확률값들을 합산하면 1.00이 된다.

3) 컴퓨터 언어 처리분석을 이용한 외교전문의 감정분석

현재까지 알려진 감정적 방향성(valence)은 대개 일차원 공간(one-dimensional space)을 가정한 것이다. 다시 말해 감정분석은 문서, 혹은 문서 내의 조각들(문단, 혹은 문장 등)의 감정적 방향성이 “긍정적·부정적”인가를 정량화시키는 것이 대부분이다. 물론 인간의 감정이 일차원이 아닌 다차원적(multidimensional) 공간일 수도 있다는 많은 연구들이 존재하지만, 오스굿 등(Osgood, Suci, & Tannenbaum, 1957)의 의미분화척도(semantic differential scale) 이래로 많은 연구들이 보여 주듯이 가장 중심적인 감정은 바로 “긍정적·부정적” 감정이다(이 문제에 대한 가장 최근의 문헌정리는 Keltner & Lerner, 2010을 참조하라).

현재 사회과학분야에서 영문서를 분석할 때 가장 널리 쓰이는 감정 분석 처리기법은 텍사스 대학(University of Texas)의 사회심리학 교수인 펜네베이커와 동료들의 LIWC[®](Linguistic Inquiry and Word Count)이다(Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007). LIWC는 프로그램 개발자들이 수십 년에 걸쳐 축적한 단어 구별 사전(dictionary)을 바탕으로 문서 속에 있는 단어들을 파악한 후 단어들의 심리학적 관점에서 의미 있다고 판단되는 범주에 해당되는 단어들의 빈도를 계산하는 아주 간단한 원리를 따른다. 그러나 단순한 원리에도 불구하고, LIWC는 문서를 작성한 사람의 심리적 변화를 매우 잘 예측하는, 즉 높은 예측 타당도(predictive validity)를 갖고 있는 언어분석 프로그램으로 알려져 있다(이귀혜 · 강남준 · 이종영, 2008⁸; Bantum &

Owen, 2009; Pennebaker, 1997; Pennebaker, Mehl, & Niederhoffer, 2003).⁹

가장 최근 버전이라고 할 수 있는 LIWC[®] 2007의 경우 80개의 언어학적·심리학적 범주들¹⁰에 해당되는 단어들의 개수를 계산해 준다. 여기서 우리는 ‘긍정적 감정어’와 ‘부정적 감정어’를 이용하여 위키리크스 외교전문에서 드러난 문서작성자의 감정을 분석하였다. 이번 연구에서는 긍정적 감정과 부정적 감정을 살펴보는 것은 물론 두 감정이 양가적(兩價的)으로 결합되어 나타나는 수준 역시 측정하였다.¹¹ 양가감정이란 긍정적 감정과 부정적 감정이 동시에 나타나는 심리적 상태를 의미한다(Thompson, Zanna, & Griffin, 1995). 양가감정을 측정하기 위해서는 사회심리학 문헌에서 가장 많이 이용되는 그리핀 공식(Griffin for-

8 이귀혜·강남준·이종영(2008)은 LIWC의 분석기법을 한국어 상황에 적용하였는데, 이들의 연구는 한국어를 대상으로 한다는 점에서 미국대사관 근무자에 의해 작성된 영어를 분석한 본 연구의 맥락과는 약간 다르다.

9 그러나 LIWC 역시도 ‘외교문서’와 같은 특화된 장르의 언어를 분석하는 데 적절하지 않을 가능성을 배제할 수 없다. 저자가 인지하는 한, 분석 대상 텍스트가 상대적으로 넓은 LIWC과는 다른 전통에서 개발된 유사한 언어분석 프로그램으로는 로더릭 하트(Hart, 1984, 2000)가 개발한 딕션(DICTION)을 들 수 있다. 그러나 딕션의 경우 정치인의 레토릭을 분석하기 위해 특화된 프로그램이라는 점에서 외교전문을 다루는 이번 연구와는 맥락이 상당히 다르다. 특화된 영역의 텍스트를 분석하기 위한 프로그램을 다른 영역의 텍스트를 분석하는 데 적용하는 경우 분석의 타당도는 매우 떨어질 수밖에 없다(Krippendorff, 2004, pp. 260~261을 보라). 프로그램이 잘 분석할 수 있는 장르의 언어가 분석대상이 되는 장르의 언어와 어긋날 때 컴퓨터 언어 처리 결과의 타당성은 매우 낮아지게 된다. 이는 본 논문의 명확한 한계점이라고 할 수 있다.

10 이들 80개의 범주들은 위계적으로 구성되어 있다. 이를테면, 이번 연구에서 주로 이용된 부정적 감정어(negative emotion)의 경우는 세부적으로 불안(anxiety), 분노(anger), 슬픔(sad)이라는 세 개의 하부 범주들을 포괄한다. 80개의 범주에 대한 정의와 예시 단어들에 대해서는 펜네베이커 등이 개발한 프로그램의 매뉴얼을 참조하라(Pennebaker, et al., 2007). 아쉽게도 이들 단어들에 대한 전체 목록은 저작권으로 보호받고 있기 때문에 본 논문에서 공표하기 어렵다. 대신 프로그램 저작권 소유자인 텍사스 대학의 펜네베이커 교수에게 문의하거나(전자 메일: jwpennebaker@gmail.com) 해당 프로그램을 이용하면 해당 단어 목록을 얻을 수 있다.

11 부정적 감정어와 긍정적 감정어 사이의 상관계수는 $r = .33$ ($p < .001$)이었다. 즉 한편의 감정이 강하게 드러난 문서에서는 다른 한편의 감정도 어느 정도 드러나고 있다고 볼 수 있다. 이러한 경험적 결과는 우리 연구에서의 양가감정 분석의 타당성을 뒷받침하고 있다.

mula)을 이용하였다(Thompson, et al., 1995). 구체적으로 그리핀 공식은 아래와 같다.

$$\text{양가감정} = \frac{\text{긍정적 감정} + \text{부정적 감정}}{2} - |\text{긍정적 감정} - \text{부정적 감정}|$$

공식에 대해서 조금 설명하자면, 우변은 2개의 부분으로 나누어진 다. 첫 번째 부분은 감정의 방향성을 고려하지 않은 채 감정의 ‘강도(intensity)’만을 정량화시킨 것이다. 두 번째 부분은 발산된 감정이 얼마나 한쪽(긍정적 혹은 부정적)으로 치우쳐져 있는가를 정량화시킨 것이다(톰슨 등(Thompson, et al., 1995)은 이 부분을 ‘유사성(similarity)’이라고 명시하였다). 이를테면 어떤 문서가 긍정적 감정과 부정적 감정의 점수의 범위가 각각 [0, 50]이라고 가정하자. 만약 어떤 문서가 긍정적 감정과 부정적 감정 모두에서 50점씩을 기록하였다면, 양가감정 점수는 50이 된다. 그러나 긍정적 감정이 0이며 부정적 감정이 50일 경우 해당 문서의 양가감정 점수는 -25가 된다. 반대로 긍정적 감정과 부정적 감정이 모두 0일 경우는 0이 된다. 즉 양가감정의 점수가 높을수록 긍정적 감정과 부정적 감정이 뒤섞여 있는 문서인 반면, 양가감정의 점수가 낮을수록 특정한 방향의 감정이 강하게 드러난 문서로 볼 수 있다.

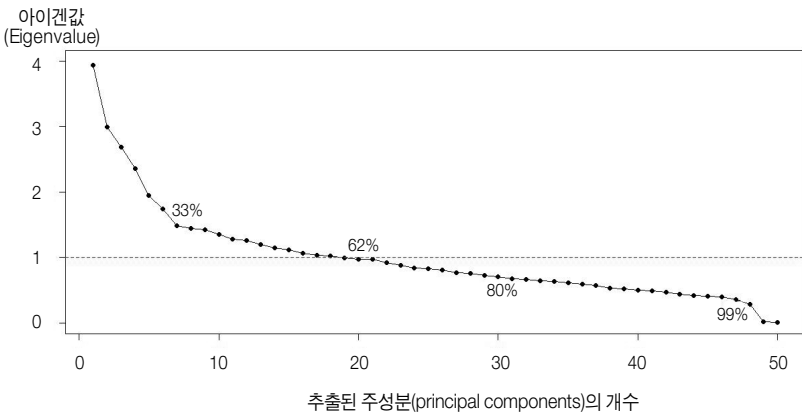
4) 측정방식 및 검증모형의 구성

분석 대상인 외교전문들은 2006년 1월 9일부터 2010년 2월 25일까지의 1,080개가 있다. 우리 연구에서는 개별 전문을 분석단위(unit-of-analysis)로 간주하였다. 분석을 위한 변수들로는 전문에서 나타난 주제의 분포, 그리고 전문에서 드러난 감정적 방향성(긍정-부정)을 주요 결과변수로, 그리고 한국과 미국의 정권변화에 따른 세 가지 시기들을 주요 예측변수로 이용하였다. 가장 핵심이 되는 예측변수는 아래처럼 3수준의 명목변수로 구성하였다.

1. “대등적－일방적” 시기: 노무현 전 대통령 그리고 부시 전 대통령 집권기
2. “협력적－일방적” 시기: 이명박 전 대통령 그리고 부시 전 대통령 집권기
3. “협력적－상호적” 시기: 이명박 전 대통령 그리고 오바마 현 대통령 집권기

수많은 구체적인 ‘단어’로부터 통제가능한 개수의 추상적인 ‘주제’를 끌어내기 위하여 우리는 2단계에 걸쳐 텍스트 자료를 처리하였다. 제 1단계는 3-2)의 끝부분에서 설명한 LDA 분석을 통하여 50개의 주제, 즉 기계적 단어들의 묶음을 추렸다. 이후 제 2단계에서는 이렇게 축약된 50개의 단어들 묶음을 주성분 분석을 실시하였다. 주성분 분석은 추출된 50개의 주제에 대한 연구자의 주관적 판단을 최소화시키며, 복잡한 현상을 간명하게 이해할 수 있도록 데이터를 축약시키기 위함이다.

그림 3. LDA를 이용해 추출된 50개의 주제들의 주성분 분석 스크리도표



알림. 가로 점선은 일반적인 분할기준(cutoff criterion)인 카이저규칙, 즉 아이겐값이 1인 경우를 표시함. 그림에 포함된 %는 주성분의 개수별 누적설명분산을 의미한다.

주성분 분석을 통해 적절한 주제들의 개수를 선택하기 위한 스크리도표 (screeplot) 결과가 <그림 3>이다. 결과에서 볼 수 있듯, 7번째 주성분부터 아이겐값(eigen-value)의 감소 속도가 완만해진다는 것을 알 수 있으며, 7개의 주성분으로 설명되는 총 분산비율은 33%였다. 또한 일반적으로 이용되는 카이저규칙(Kaiser rule), 즉 아이겐값이 1.00을 넘는 주성분의 개수는 총 19개이며, 이들을 모두 사용할 경우 설명되는 총 분산 비율은 62%였다.

일반적 기준, 즉 카이저규칙을 따를 경우 고려하여야 할 주성분은 총 19개이지만, 여기서는 간명함(parsimony) 원칙을 중시하여 7개의 주성분만을 연구에서 고려하였다. 물론 19개의 주성분들을 선택할 경우 29%의 추가분산이 설명되지만, 고려하여야 할 잠재주제들의 숫자가 거의 3배로 늘어난다는 부담이 있다. 또한 <그림 3>의 스크리도표에서 볼 수 있듯 7개를 기준으로 구분할 경우, 하나의 주성분 당 평균 설명분산은 4.7%인 반면, 8번째부터 19번째까지의 주성분의 경우 평균 설명분산이 2.2%로 상당한 차이가 나타났다. 실제로 이는 아이겐값이 1 미만인 20번째에서 30번째 주성분들의 평균설명분산인 1.9%와 그다지 많은 차이를 보이지 않고 있다. 이러한 점에서 7개의 주성분만을 선택하는 것에는 무리가 없다고 판단하였다. 7개의 주성분은 ① 한국 내 정세, ② 북한 관련, ③ 주한미군, ④ 남북경협, ⑤ 해외파병, ⑥ 동북아문제, ⑦ 한미 통상이었다.¹²

결과변수의 또 다른 한 갈래인 외교전문에서 나타난 감정적 표현의 경우 LIWC[®] 2007 (Pennebaker, et al., 2007)을 통하여 분석하였다. 앞에서 설명하였듯이 LIWC[®] 2007의 단어분류를 따라 각 외교전문에서 긍정적 감정을 나타내는 단어들의 비율과 부정적 감정을 나타내는 단어들의 비율을 각각 계산하였다($M_{\text{긍정}} = 1.36, SD_{\text{긍정}} = .49; M_{\text{부정}} = .69, SD_{\text{부정}} = .39$).

12 주성분분석의 구체적 결과는 저자에게 문의하면 얻을 수 있다.

또한 개별 외교전문에 등장한 긍정적 감정과 부정적 감정 단어들의 비율을 앞서 설명한 ‘그리핀 공식’에 기반하여 양가감정(兩價感情)을 계산하였다($M=0.01$, $SD=.49$).

또한 결과변수와 예측변수의 관계에 영향을 미칠 수 있는 변수들을 통제하였다. 우선 각 외교전문들에 이용된 단어들의 개수, 즉 ‘문서의 분량’을 통제하였다. 상식적으로 문서가 길수록 많은 단어를 사용하게 되며 또한 상세한 설명을 하게 되는 경우가 많다. 일반적인 외교전문 길이는 약 2,417 단어였다($SD=1714.73$). 즉 개별 문서들의 길이는 주제의 등장가능성이나 감정의 표출에 잠재적으로 영향을 줄 수 있다. 두 번째로 문서의 보안등급을 이분(dichotomous) 변수로 전환하여 통제하였다. 언급하였듯이 ‘비분류일반’의 경우 분석대상에서 제외되었기 때문에, 문서의 보안등급은 ‘대외비’를 0으로 ‘비밀’을 1로 변수화하였다($M=.11$, $SD=.36$). 마지막으로 ‘인지적 과정(cognitive mechanism)’을 의미하는 단어들의 비율을 통제하였다. 인지적 과정이란 since, because, cause 등의 단어로 측정되는데, 펜네베이커 등(Pennebaker, et al., 2007)에 따르면 이러한 단어들로 문서 저자의 인지적 사고수준을 측정할 수 있다고 한다(Pennebaker, et al., 2003도 참조). 외교문서와 같은 공식적 문서에서의 감정표출은 “~이기 때문에, ~하다” 등과 같은 형태로 나타나기가 쉽다는 점에서 인지적 과정이 두드러진 글쓰기는 보다 풍부한 감정적 표현으로 나타날 가능성이 높다고 하겠다. 문서전반에 걸쳐 약 5.42% 정도가 인지적 과정을 표상하는 단어들이므로 나타났다($SD=2.00$).

5) 가설검증을 위한 통계적 방법

구체적인 통계적 유의도 검증에 앞서서 우리는 외교전문 작성일이 시계열(time-series)이라는 점을 감안하여 자료의 자기상관성을 먼저 고려하였다. 즉 우리가 분석의 주제로 삼은 시기가 아니라 시간에 따라 결과변수에서 백색잡음(white noise)이 보이는지를 먼저 살펴보았다.¹³

시기에 따른 정규적 트렌드는 발견되지 않았으며, 따라서 일반화선형모형(Generalized Linear Model)을 적용하였다. 전술한 측정치를 이용하여 ‘대등적-일방적’, ‘협력적-일방적’, ‘협력적-상호적’의 세 시기별로 추출된 7개의 최종주제들의 발현가능성과 감정적 표현 형태가 어떻게 달라지는지 살펴보았다.

또한 언급하였듯이 문서의 길이, 보안등급, 그리고 인지적 과정 관련 단어비율을 각 모형별로 통제하였다. 또한 감정적 표현이 문서가 다루는 주제에 따라서 달라질 수도 있다는 점을 고려하여 감정적 표현을 점검하는 경우 7개의 최종주제들을 추가적으로 통제하였다.

4. 연구결과

1) 3가지 시기별 외교전문의 주제변화

첫 번째 연구문제는 “세 시기에 걸쳐 외교전문에 실린 주제들은 어떻게 변화하고 있는가?”였다. 추출된 7가지의 주제들, 즉 ① 한국 내 정세, ② 북한관련, ③ 주한미군, ④ 남북경협, ⑤ 해외파병, ⑥ 동북아문제, 그리고 ⑦ 한미통상 등을 대표하는 단어들의 발견가능성을 결과변수로 선택한 일반화 선형모형의 분석결과는 <표 2>와 같다. 또한 각 시기별 7가지 주제들이 어떻게 다르게 등장했는가를 시각적으로 분석한 결과는 <그림 4>와 같다.

<표 2>의 결과에서 3가지 주제, 즉 한국 내 정세, 남북경협, 그리고

13 시계열자료에서는 흔히 계절성(seasonality), 즉 요일이나 월에 따른 규칙적 패턴이 나타나기도 한다. 만약 시계열자료가 백색잡음(white noise)의 형태를 띠지 않으면, ‘트렌트 제거(de-trending)’를 통해 백색화(whitening) 과정을 밟는다(Wei, 1994). 비록 본문에는 보고되지 않았으나, 자기상관함수(auto-correlation function)와 부분자기상관함수(partial auto-correlation function)를 점검하였으며, 시계열변화는 백색잡음에 가까운 것을 확인하였다. 저자에게 요청하면 구체적 결과를 얻을 수 있다.

표 2. 한국-미국의 정권교체에 따른 세 시기별 7가지 주제들의 발견가능성 비교

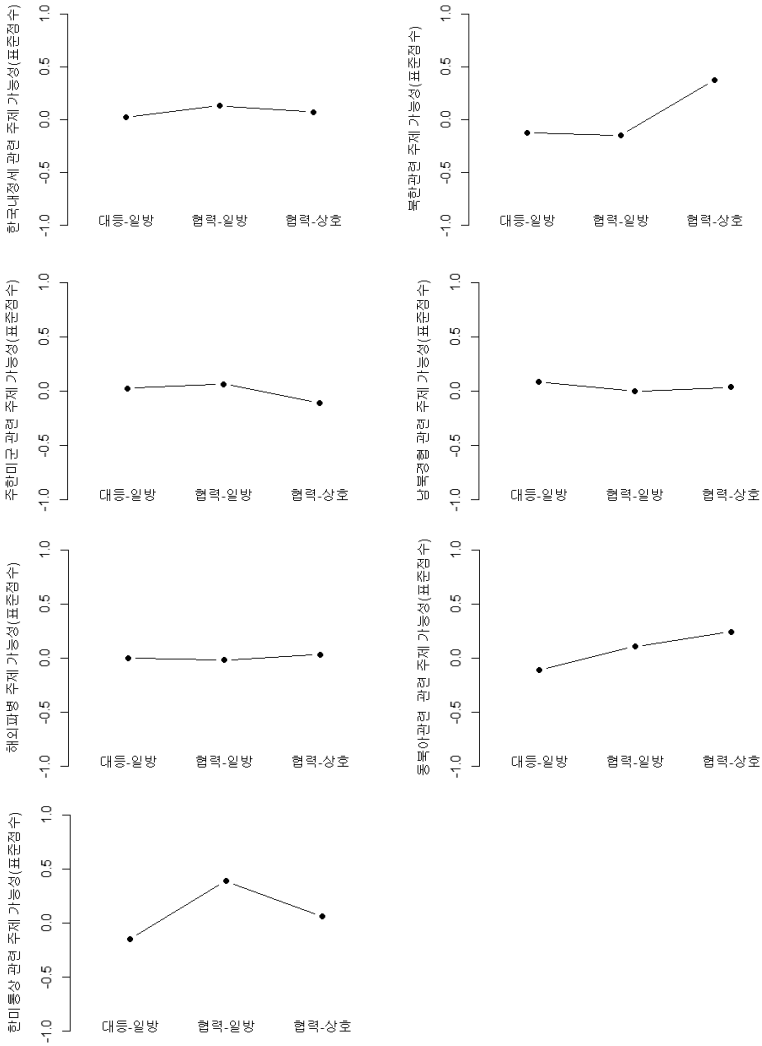
	한국 내 정세	북한 관련	주한 미군	남북 경협	해외 파병	동북아 문제	한미 통상
유의도검정							
시기구분($df=2$)	0,91	26,67***	7,57***	1,54	0,10	11,63***	26,28***
비밀구분($df=1$)	63,19***	5,51*	22,58***	69,99***	2,65	11,90***	14,01***
문서길이($df=1$)	50,13***	36,23***	209,42***	87,87***	1,66	7,39**	12,68***
인지적과정($df=1$)	7,83**	1,91	75,88***	15,59***	8,12***	8,08**	23,05***
조정된 평균							
대등적-일방적	0,025 ^A	-0,124 ^A	0,030 ^A	0,086 ^A	0,004 ^A	-0,108 ^A	-0,145 ^A
협력적-일방적	0,134 ^A	-0,144 ^A	0,069 ^B	0,001 ^A	-0,015 ^A	0,112 ^B	0,392 ^C
협력적-상호적	0,075 ^A	0,380 ^B	-0,103 ^C	0,040 ^A	0,037 ^A	0,246 ^C	0,063 ^B

알림. 유의도 검정의 경우 F 값을 표시하였으며, 잔차(residual)의 df 는 1075이다($N=1080$). 조정된 평균의 경우 비밀구분, 문서길이, 인지적과정관련 단어비율의 세 가지 통제변수들의 효과를 조정한 후 시기별로 각 주제들의 발견가능성(표준점수)을 보고한 것이다. 조정된 평균의 경우, 위치자가 공유되지 않을 경우 봉페로니(Bonferroni)의 수정된 통계적 유의도($p < .05$)에서 평균들이 통계적으로 유의미하게 차이를 의미한다.

해외파병의 경우 세 시기별로 발견가능성에 통계적으로 유의미한 차이가 없었다. 반면 북한관련, 주한미군, 동북아문제, 그리고 한미통상의 경우 각각의 주제들이 외교전문에서 발견될 가능성은 세시기별로 상이하게 나타남을 알 수 있었다.

우선 통계적으로 유의하지 않은 결과를 살펴보자. 그러나 각 시기별 해당 주제들의 발견가능성 평균값은 비교적 각 시기의 한국-미국 간의 외교적 현안을 잘 보여주고 있다. 이를 테면 한국 내 정세의 경우 대등적-일방적 기간에 가장 높은 값을 보여주었는데, 실제로 이 시기의 외교전문들은 한국의 지난 대통령 선거관련 동향을 분석, 보고한 것들이 많았다. 남북경협의 경우 대등적-일방적 기간에 가장 높고, 협력적-상호적 기간이 그 다음을 차지하고 있었다. 실제로 노무현 전 대통령 시기는 남북경협에 상당한 관심을 쏟았던 기간이었으며, 오바마 현

그림 4. 각 시기별 7가지 주제의 발견가능성 비교



알림. <표 2>에 제시된 모형에 기반하여 각 결과변수의 예측값을 도식(圖示)함.

대통령의 당선 이후 북한에 대한 식량문제 등이 언급되었던 것을 생각하면 통계적으로 유의하지 않은 차이지만, LDA-주성분분석을 통한 7가지 주제의 추출이 상당한 타당성을 보이고 있음을 알 수 있다.

LDA-주성분분석을 통한 분석의 타당성을 보다 강하게 엿볼 수 있는 발견은 바로 통계적으로 유의미한 차이가 드러난 북한관련, 주한미군, 동북아문제, 그리고 한미통상의 4가지 주제들이었다. 북한관련 언급은 협력적-상호적 시기에 가장 높게 발견되었는데, 이는 오바마 행정부의 식량지원 움직임과 밀접하게 관련되어 있다. 또한 중국, 일본, 아시아 등의 단어로 대표될 수 있는 동북아 문제 역시 시간이 지나면서 증가하는 추세를 보인다. 주한미군 관련 언급의 경우 오바마 행정부가 들어서면서 등장하는 빈도가 상대적으로 감소되고 있다. 군사력이 일방주의 정책을 지탱하는 가장 중요한 요인이며, 다자간의 협상이 상호주의 정책의 핵심적 요소임을 감안할 때, 이러한 발견은 이해될 수 있다.

특히 흥미로운 것은 한미통상문제, 보다 구체적으로 FTA나 소고기 등의 이슈들은 협력적-일방적 시기에 집중적으로 나타나고 있다는 점이다. 특히 촛불집회와 관련하여 미국대사관이 통상문제에 관한 한국의 동향을 집중적으로 보고하고 있었음을 알 수 있는 대목이다.

위와 같은 발견에 기반하여, 우리는 다음과 같은 소결론을 얻을 수 있다. 첫째, 분석결과는 LDA와 주성분 분석을 2단계로 실시한 우리의 분석이 어느 정도의 타당성을 확보하고 있다는 점을 명확하게 보여준다. 둘째, 대등적-일방적 기간에는 상대적으로 남북협력에 관한 주제들이 많았지만, 그 외로는 많이 두드러지게 드러나는 주제가 관찰되지 않았다. 셋째, 협력적-일방적 기간에는 미국의 한국 내 정세를 보고하며 무엇보다 한미통상 관련 주제들을 집중적으로 다루었다는 점이다. 즉 미국은 외교대상국의 정권교체에 민감하게 반응한다고 볼 수 있다. 넷째, 협력적-상호적 기간에는 주로 외교관련 이슈들이, 즉 동북아정세나 북한문제와 같은 이슈들이 집중적으로 등장하였다는 점이다.¹⁴

2) 3가지 시기별 외교전문외 감정적 표현변화

두 번째 연구문제는 “세 시기에 걸쳐 외교전문에 실린 감정적 표현들은 어떻게 변화하고 있는가?”였다. 언급하였듯, LIWC^① 2007의 단어분류 사전에 근거하여 각 외교전문에서의 긍정적 단어 비율과 부정적 단어들의 비율, 그리고 이 두 측정치를 이용하여 양가감정을 산출하였다. 세 가지의 감정적 표현들을 결과변수로 선택한 일반화 선형모형의 분석결과는 <표 3>과 같다. 분석에 투입된 예측변수와 통제변수들은 주제들의 변화분석과 같지만, 특정한 주제에 따라 감정적 표현이 서로 다를 수 있기 때문에, 각 주제들의 발견가능성을 추가적으로 통제한 결과도 보고하였다. 또한 각 시기별로 세 가지의 감정적 표현들의 변화패턴을 시각적으로 도시한 결과는 <그림 5>에 실려 있다.

협력적-상호적 시기에는 긍정적 표현들이 약간 감소한 반면, 부정적 표현들이 상승하는 모습을 확인할 수 있었다. 특히 흥미로운 점은 긍정적 감정과 부정적 감정이 공존하는 심리적 상태인 양가감정이 협력적-상호적 시기에 매우 높아졌다는 점이다. 더욱이 흥미로운 기간은 대등적-일방적 시기다. <그림 5>의 왼쪽과 가운데 그림에서 볼 수 있듯, 평균적인 긍정적 감정, 혹은 부정적 감정은 딱히 높지도 혹은 낮지도 않음을 알 수 있다. 그러나 개별 외교전문에 집중할 경우 긍정적 감정이나 부정적 감정 하나의 감정만이 강하게 나타나는 시기가 바로 대등적-일방적 시기다. 이러한 발견은 이는 주제들의 변화를 통제한 후에도 발견된 것이라는 점에서 단순히 주제의 변화에 따라 감정이 달라졌다고 치

14 여기서 발견된 결과는 그 자체로 상식에 어긋나거나 혹은 놀라운 발견은 없다. 이런 점에서 과연 이 논문의 현실적 의의가 별로 없다는 지적도 어느 정도는 타당하다. 그러나 지적의 타당성에도 불구하고 상식에 부합하는 연구결과 역시 충분한 연구의의를 갖는다고 믿는다. 그 이유는 두 가지다. 첫째, 경험적 분석을 통해서 상식을 확인(confirm)하는 것 역시 중요한 과학적 업 중 하나다. 새로움(novelty)을 추구하는 것 못지않게 상식적인 것을 확인함으로써 상식의 범위를 확정짓는 작업은 그 자체로 충분한 의의가 있다고 믿는다. 둘째, 새로운 것 혹은 상식에 어긋나는 것만을 발표할 경우 출판편향(publication bias)에 빠질 수도 있다.

표 3. 한국-미국의 정권교체에 따른 세 시기별 감정적 표현들의 수준 비교

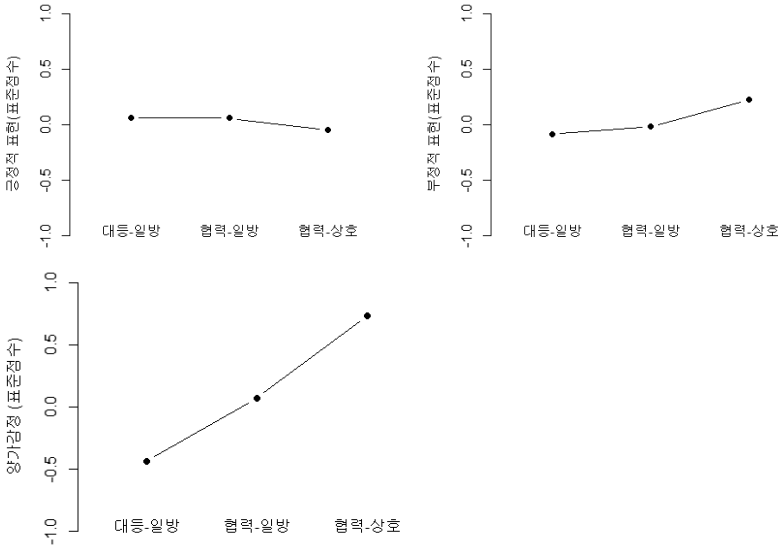
	긍정적 표현		부정적 표현		양가감정	
유의도검정						
시기구분(df=2)	5.97**	8.06***	9.34***	10.93***	177.68***	177.58***
비밀구분(df=1)	96.92***	130.80***	58.62***	68.60***	0.47	0.47
문서길이(df=1)	122.69***	165.58***	157.92***	184.81***	0.05	0.05
인지적과정(df=1)	476.97***	643.70***	217.49***	317.73***	0.39	0.39
한국 내 정세(df=1)		122.52***		74.62***		0.33
북한관련(df=1)		79.80***		48.86***		2.78
주한미군(df=1)		128.53***		5.74*		2.74
남북경협(df=1)		0.62		1.55		0.06
해외파병(df=1)		28.85***		0.21		0.33
동북아문제(df=1)		0.34		10.92***		0.04
한미통상(df=1)		21.40***		47.87***		0.14
조정된 평균						
대등적 - 일방적		0.064 ^A		-0.080 ^A		-0.432 ^A
협력적 - 일방적		0.062 ^A		-0.014 ^A		0.072 ^B
협력적 - 상호적		-0.047 ^B		0.228 ^B		0.734 ^C

알림. 유의도 검정의 경우 F 값을 표시하였으며, 잔차(residual)의 df 는 1068이다($N=1080$). 조정된 평균의 경우 비밀구분, 문서길이, 인지적 과정 관련 단어비율의 3가지 통제변수들의 효과와 추가적으로 LDA-주성분분석을 통해서 추출된 7가지의 주제들의 발견가능성을 조정한 후, 시기별로 각 주제들의 발견가능성(표준점수)을 보고한 것이다. 조정된 평균의 경우, 위첨자가 공유되지 않을 경우 봉페로니(Bonferroni)의 수정된 통계적 유의도($p < .05$)에서 평균들이 통계적으로 유의미하게 차이를 의미한다.

부하기 어렵다는 점을 감안하여야 한다.

그렇다면 위와 같은 발견이 뜻하는 것은 무엇인가? 분석의 측면에서 우리의 분석이 주제별 감정분석을 하지 않고, 문서별 감정분석을 하였기 때문에 나타나는 문제를 배제할 수 없다. 현재의 감정분석으로는 ‘주제1 - 긍정’, ‘주제2 - 부정’을 포함하는 문서의 경우, 주제별로는 감정적 표현이 명쾌하며 전혀 양가감정이 없으나, 두 가지의 주제를 구분하

그림 5. 각 시기별 감정적 표현들의 변화패턴



알림. <표 3>에 제시된 모형에 기반하여 각 결과변수의 예측값을 도식(圖示)함.

지 않을 경우 해당문서에 양가감정이 나타나는 것으로 잘못 판단하게 되는 문제점이 있을 수 있다. 그러나 대부분의 외교전문들을 살펴보면, 해당되는 주제가 하나인 상태에서 발송되며 만약 2개의 주제가 하루에 발송되는 경우 2개의 개별적 외교전문을 보내는 경우가 대부분이었다. 즉 보다 타당한 해석은 우리는 대등적-일방적 시기의 외교전문들은 긍정적 평가이거나 그렇지 않으면 부정적 평가를 보인 반면, 협력적-상호적 시기에는 두 가지 평가를 하나의 외교전문에서 동시에 나타나기가 쉽다는 점이다. 특히 이는 이론적 논의 부분에서 언급하였듯, 한국의 북한 이슈, 혹은 동북아관련 이슈가 포함되면 상호주의적 외교정책을 추진하려는 미국정부의 입장에서 대미외교를 중시하는 현 한국정부가 좋을 수도 혹은 걸끄러울 수도 있음을 암시한다고 하겠다. 실제로 이전의 <표 3>과 <그림 4>에서도 나타나듯 협력적-상호적 시기에는 외교관련 이슈

들이 집중적으로 등장하고 있으며, 또한 해당 이슈들을 다루는 세 번째 시기에 양가감정이 가장 늘어난다는 것도 우리의 해석을 강하게 뒷받침하고 있다.

예를 들어 이명박 후보자가 대통령으로 당선된 후에 작성된 2008년 1월 13일 외교전문(문서번호 08SEOUL67, 대외비)에는 노무현 정권의 이재정 통일부 장관의 3자회담관련 발언 및 이명박 대통령이 통일부를 축소할 것이라는 전망을 전하면서 불편한 심기를 드러내고 있었다. 특히 해당 외교전문의 경우 Although나 While 등의 접속사를 많이 쓰고 있었다. 반면 오바마 대통령 취임 이후 게이즈 국방장관의 방한 당시 작성된 외교전문에서는 현 한국 정부에 대한 긍정적 평가와 이에 대한 북한의 부정적 평가, 그리고 당시 공전중이었던 국회에서의 문제 등을 다루었다(이를테면 2009년 10월 15일, 문서번호 09SEOUL1631, 비밀).

5. 결론 및 연구의 함의

위키리크스에서 미국대사관에서 본국으로 발송한 외교전문들이 공개된 후, 미국 정부는 외교전문을 공개했던 매닝(Manning) 일병과 폭로를 주도했던 어산지(Assange)를 구속할 정도로 강하게 반발하였다. 위키리크스의 외교전문 공개행위는 사회적, 정치적 논란을 불러왔으며, 다양한 시각을 갖는 학술적 연구활동으로 이어졌다. 그러나 뜨거운 논란의 대상이 되었지만, 의외로 위키리크스에서 공개된 내용이 무엇인가에 대해 포괄적인 경험적 분석을 시도한 연구는 없었다. 대부분의 연구들은 위키리크스가 불러온 파장에 대해 '표현의 자유'와 '안보' 사이의 바람직한 관계형성을 탐구하거나 위키리크스의 정보공개행위가 저널리즘적 관점에서 어떤 의미를 갖는지를 규범적으로 논의하거나 혹은 몇몇 주목할 만한 사례들(anecdotes)을 근거로 위키리크스 문서에 대한 총체적인 평가(이를테면 권위주의 국가의 민주화에 기여함)를 내리는 것이 대부분이었다.

그러나 이번 연구에서는 위키리크스 그 자체에 어떤 내용이 실려져 있는가를 컴퓨터 언어처리 기법을 이용하여 탐색하였다. 구체적으로 이번 연구는 외교적 커뮤니케이션 과정 속에서 대사관이 갖고 있는 매개자적 속성, 즉 본국과 이해관계와 관점을 공유하고 주재국과는 시·공간적 경험을 공유하는 대사관의 커뮤니케이션 역할에 초점을 맞추었다. 흥미롭게도 위키리크스에서 공개된 문서는 한국과 미국의 정권교체 기간과 맞물려 있어, 대사관의 커뮤니케이션적 특징이 주재국과 본국의 정치환경 변화에 의해서 어떻게 달라지는가를 살펴볼 수 있는 좋은 환경을 제공하고 있었다. 한국과 미국의 정권교체에 따른 대사관 커뮤니케이션 특징을 살펴보기 위해서 우리는 ‘노무현 전 대통령-부시 전 대통령’, ‘이명박 전 대통령-부시 전 대통령’, 그리고 ‘이명박 전 대통령-오바마 현 대통령’의 시기를 구분하여 각각 ‘대등적-일방적’, ‘협력적-일방적’, 그리고 ‘협력적-상호적’라고 개념화하였다. 또한 방대한 분량의 위키리크스 문서들은 최근에 개발된 자연어 처리기법 중 하나인 ‘잠재적 디리실레분포 할당모형(LDA, Latent Dirichlet Allocation)’을 이용하여 문서의 50개 주제들을 기계적으로 추출한 후, 50개의 주제들을 주성분 분석을 통해 축약하여 가장 두드러지는 7개의 최종주제들을 확정하였다. 또한 문서에 드러난 대사관의 감정적 표현을 살펴보기 위해 사회심리학에서 타당성을 인정받은 LIWC® 2007을 이용하여 긍정적 표현과 부정적 표현, 그리고 문서에서 드러난 양가감정(兩價感情, 긍정적 감정과 부정적 감정의 공존 정도)을 측정하였다.

3가지 시기별로 문서에서 드러난 주제들과 감정적 표현을 분석한 결과 다음과 같은 사항들을 확인할 수 있었다. 첫째, LDA와 주성분 분석을 통하여 기계적으로 추출된 언어분석은 일정수준 이상의 타당성을 보여주었다. 물론 텍스트의 의미를 가장 타당하게 해석할 수 있는 존재는 오직 인간뿐이지만(Krippendorff, 2004, p. 259), 인간이라면 누구나 갖고 있는 주관적 편향과 시간과 비용, 그리고 능력의 한계라는 물리적 제약으로 대용량의 텍스트를 컴퓨터를 이용하여 기계적으로 처리하는 것은 높은 신뢰도와 효율성을 가질 수 있다는 장점이 있다(Krippendorff,

2004, 제12장). 이번 연구에서는 컴퓨터를 이용한 텍스트 분석이 일정 수준의 타당성을 확보하고 있음을 보여준다는 면에서 큰 의의가 있다고 할 수 있다.

둘째, 외교전문에 등장하는 주제는 주재국보다는 본국의 정권교체에 더욱 민감하게 반응하여 변화하고 있었다. 물론 남북경협과 같은 경우는 주재국과 보다 밀접하게 연관된 이슈이기는 하지만, 이 경우 눈에 띄는 패턴은 발견할 수 없었다. 보다 두드러지는 차이는 부시 행정부에서 오바마 행정부로 바뀌면서 보다 두드러지는 주제들이었다. 이러한 차이는 일종의 선택적 주의(selective attention)의 일종으로 볼 수 있다. 다시 말해 관측자의 선유경향(predisposition)에 따라 현상의 특정부분을 집중하여 주목하는 '선택적 주의'처럼, <그림 1>과 같이 대사관은 본국의 시각으로 주재국을 바라보고 있다. 이는 대외정책의 기초자료로 쓰이는 외교문서에 나타난 주재국은 본국의 시각으로 구성된 현실(constructed reality)임을 암시하고 있다. 이 같은 선택적 주의가 발생한 이유는 주재국의 현 상황 혹은 변화를 기술할 때, 어떠한 부분을 부각시킬 것인가는 본국의 이해관계 혹은 관점에 기반하기 쉽기 때문일 것이다. 즉 한국의 현실 중 미국과 관계가 되는 부분들이 언급되며, 이러한 경향이 위키리크스에서 공개된 외교전문에서도 드러난다고 볼 수 있다. 약소국의 입장에서 강대국이 행사하는 선택적 주의는 독자적 외교정책의 현실적 한계를 암시하고 있다. 즉 약소국이 스스로를 바꾸려고 노력해도, 강대국이 바뀌지 않는 한 '강대국이 바라본 약소국'은 변하지 않는다. 다시 말해 약소국의 독자적 외교정책 변화는 강대국의 정책결정 자료인 외교문서에 제대로 반영되지 않으면서 결국 실현의 가능성은 더욱 줄어들게 된다. 펠로폰네소스 전쟁사에서 투키디데스가 이야기한 이후 국제관계학의 주요 이론으로 정립된 고전적 현실주의(classic realism)는 어쩌면 강대국의 외교관이 구성한 주재국의 모습에서 어느 정도 비롯되었을 수 있다.

셋째, 외교전문에 나타나는 감정적 표현은 본국의 외교정책 노선과 주재국 정권의 성향과 관계가 있지만, 감정적 표현의 결정적 변화는 결

국 주재국의 정권변화보다는 본국의 정권변화와 밀접하게 연관되어 있다. <표 3>와 <그림 5>에서 나타나듯, 감정적 표현의 변화는 한국의 정권변화보다는 미국의 정권변화에 더 민감하게 반응한다. 특히 협력적-상호적 시기로 대표되는 제3시기의 경우 긍정적 감정의 감소와 부정적 감정의 증가, 그리고 무엇보다 양가감정의 증가는 한국에서 대미외교와 한미관계를 중시하는 정권이 들어선다고 해도, 본국인 미국이 바뀌지 않는다면 주재국인 한국을 반드시 호의적으로 바라보지 않는다는 것을 의미한다. 이러한 점에서 외교적 정책결정자의 한국정치인 혹은 정책결정자에 대한 평가는 ‘누가 평가를 하는가?’를 중심으로 분석되고 보도되어야 한다. 우리 언론보도의 경우 전직 대통령이나 유력 정치인에 대한 평가를 보도할 때 평가하고 있는 사람에 대한 설명을 등한하게 하는 경향이 있다. 특히 최근 조지 W. 부시나 콘돌리자 라이스, 로버트 게이츠 등의 회고록에 묘사된 전직 대통령에 대한 평가를 ‘누가 평가했는가?’에 대한 맥락 정보 없이 보도하는 경우가 많다. 전직 대통령에 대한 평가는 양극화된 국내정치 현실과 맞물려 불필요한 사회적 갈등을 야기하기도 한다. 하지만 이번 연구에서 암시하듯 ‘누구를 어떻게 평가했는가?’는 그다지 중요하지 않을 수 있다. 대신 ‘누가 어떤 시각에서 평가했는가?’를 보도한다면, 독자가 외교관련 국가 간 커뮤니케이션을 보다 더 타당하게 이해할 수 있을 것이다.

위와 같은 발견과 컴퓨터 이용 텍스트 분석의 함의에도 불구하고, 이번 연구는 다음과 같은 몇 가지 한계점을 노정하고 있다. 첫째, 분석되는 텍스트가 외교문서라는 점에서, 중요한 뉘앙스를 놓칠 수 있는 개연성이 높다. 외교문서를 포함하여 정무문서의 경우는 대부분 형식화된 포맷으로 제작되며, 문서의 작성자 역시 가능하면 감정적 표현을 자제하며 건조하게 현상을 묘사하려는 경향이 있다. 특히 대사관의 전문들은 본국의 의사결정자들에게 참고자료의 제공을 그 목적으로 하기 때문에, 감정적 표현을 해석할 때 크게 주의할 필요가 있다. 텍스트의 질적인 특성을 파악하기 위한 가장 좋은 방법은 텍스트에 대한 일정수준 이상의 지식을 갖고 있는 사람이 직접 읽고 판단하는 것이다. 따라서 향후 연구에서는 일

정수준의 지식과 언어능력을 보유하고 있는 코더를 통해 질적분석과 양적 분석을 병행하는 방식으로 연구의 타당성을 높여야 할 것이다.¹⁵

둘째, 출현빈도를 기반으로 하는 LDA나 LIWC의 접근은 출현빈도가 낮기 때문에 의미 있는 단어를 배제하기 쉽다. 분명히 a, the 등의 기능어를 제외하고 문서에서 반복적으로 나타나는 단어는 문서의 중심 주제가 될 확률이 높으며(LDA의 가정), 긍정적 단어를 많이 쓰는 경우 긍정적 감정이 보다 강하다고 볼 수 있다(LIWC의 가정). 그러나 특히 몇몇 논란이 되었던 용어들의 경우 발생빈도가 매우 낮기 때문에 큰 의미를 갖는 경우가 있다. 이를테면 앞에서 언급했던 “철두철미하게(to the core)”라는 표현의 경우 core의 빈도는 별로 높지 않지만, 잘 드러나지 않는 표현이라는 점에서 오히려 더 큰 의미와 중요성을 갖고 있다. 즉 이번 연구에서 채택한 컴퓨터 텍스트 분석의 경우 출현빈도가 낮은 중요 단어들의 의미가 희석될 위험성이 있다.

셋째, 인용된 발언의 경우 그것이 과연 대사관의 시각으로 대치될 수 있는가에 대한 의문이 있다. 이를테면 앞서 언급한 이명박 전 대통령의 친형인 이상득 국회의장의 발언을 과연 미국 대사관의 시각으로 받아들일 수 있는가 논란이 될 수 있다. ‘적의 적은 동지’라는 식으로 이어 기하하자면, 만약 ‘협력적-일방적’ 시기에 주한 미국대사관이 북한에 대

15 현재 본 논문에서 분석한 1,080개 정도의 문서를 코더를 이용해서 내용분석하는 것은 현실적으로 불가능하지는 않을지 모른다. 실제로 코더를 통한 내용분석을 시도한 결과 연구자 주변에서는 신뢰 가능한 코더를 구하는 것이 쉽지 않았다. 우선 내용분석 대상이 영문인 반면, 코더의 모국어는 영어가 아니었다. 또한 외교문서에 쓰이는 단어가 평이한 수준이 아니며 문장의 수준이나 표현방식 역시 영어가 모국어가 아닌 코더가 읽기에 쉽지 않았다. 실제로 연구자 주변의 대학원생을 대상으로 한편의 외교전문을 읽는 데 걸린 시간을 체크한 결과 하나의 외교전문당 대개 1시간 정도가 걸리는 것으로 나타났다(전문 하나는 약 2,400단어로 대략 A4기준 8페이지 내외의 길이를 가지고 있다. 흔히 해외 언론학 논문집에 실리는 논문의 본문이 6,000단어 이내(참고문헌과 도표 제외)임을 감안할 때 코더가 내용분석을 수행하면서 느끼는 피로가 상당하다고 판단하는 것이 불가능한 것은 아니다). 물론 영어 표현에 능숙하면서 한미 관계에 대해서 일정정도 지식이 있는 코더를 구하는 것도 한 가지 방법이겠지만, 이러한 코더를 구하는 것은 현 단계의 연구자에게 현실적으로 무리라고 판단하였다.

해서 부정적 발언을 한 국내 정치인을 인용하였다고 가정하자. 이 경우 문서의 내용은 북한을 규제하려는 미국의 일방주의적 외교시각에 기반할 경우 한국의 정세에 대해 긍정적 태도를 보이고 있다고 볼 수 있지만, 컴퓨터 내용분석상 해당문서는 부정적 감정을 많이 보이는 것으로 코딩될 수밖에 없다. 현 상황에서 컴퓨터 언어처리분석에 기반하여 이러한 사례들의 비율을 계산하는 것은 쉽지 않은 일이다. 그러나 대체적으로 외교전문에서의 내용과 맥을 같이 하는 것들이 인용된다는 가정을 받아들이지 않을 경우, 이 문제는 향후의 연구가 반드시 풀어야 할 숙제라고 할 수 있다.

이러한 문제점들에도 불구하고 이번 논문에서는 방법적 측면에서 그리고 외교 커뮤니케이션과 관련된 이론의 측면에서도 몇 가지 부정할 수 없는 성과들을 거두었다. 우선 방법론적 측면에서는 컴퓨터를 이용한 텍스트 분석으로도 어느 수준의 타당성은 확보할 수 있다는 것이다. 대용량의 텍스트를 효율성과 신뢰성을 보장하면서 다루기 위해서는 결국 컴퓨터에 의존할 수밖에 없는 것이 현실이다. 또한 이론적 측면에서는 미국이 바라보는 한국의 모습은 한국의 변화에 의해서 추동된다기보다 미국 내의 변화에 의해서 초래된다는 점을 확인했다는 사실이다. 물론 이와 같은 발견이 주재국의 정세변화가 아무런 의미가 없다고 말하는 것은 아니다. 단지 변화의 주요요인은 주재국보다는 본국의 정세변화라는 점이다. 특히 이번 연구 결과는 대미외교가 중시되는 한국의 입장에서 한반도를 둘러싼 국제정세 속에서 한국의 역할이 얼마나 제한적인지를 웅변하고 있다.

향후 연구에서는 미국의 대외정책에 대한 한국의 언론보도와 외교전문에 나타난 미대사관의 관점, 그리고 미국 신문에 나타난 미국의 대외정책 등을 상호비교한다면 위키리크스의 위상을 보다 정확하게 가늠할 수 있을 것이다. 또한 주일대사관이나 주중대사관 등의 발송문서 등과 상호비교함으로써 동아시아 속에서 한국이 미국에 대해서 갖고 있는 의미는 무엇인지를 살펴볼 수도 있다. 혹은 문서 속의 주제와 감정적 표현만을 살펴보았던 이번 연구의 범위를 넘어 주제어 네트워크 분석

(thematic/semantic network analysis, 예를 들어 van Atteveldt, Kleinnijenhuis, & Ruigrok, 2008 참조)을 실시하여 외교전문에서 한국은 어떻게 이해되며 어떠한 개념을 중심으로 이해되고 있는지를 살펴보는 것도 향후 과제 중 하나라고 하겠다.

참고문헌

- 강남준·이종영·오지연 (2008). 신문기사의 표절 가능성 여부 판정에 관한 연구: 컴퓨터를 활용한 형태소 매칭기법을 중심으로 『한국언론학보』, 52권 1호, 437~466.
- 길윤희 (2010. 12. 13). 메시아인가? 무분별한 철부지인가? 『한겨레 21』.
- 김두섭·강남준 (2008). 『회귀분석: 기초와 응용』. 서울: 나남
- 김유향·조희정 (2011). 튀니지의 자스민 혁명과 SNS의 역할. 『이슈와 논점』, 192호
- 김현태 (2011). 온라인 내부고발 저널리즘으로서의 위키리크스 보도에 대한 비판적 담론분석(CDA). 『커뮤니케이션학 연구: 일반』, 19권 1호, 127~15.
- 권웅 (2010). 가공할 외교테러에 외교사령탑 끄공. 『시사인』, 2010. 12. 11일 기사.
- 마상윤·박원근 (2009). 데탕트기의 한미갈등: 닉슨, 카터와 박정희. 『역사비평』, 통권 86호, 113~139.
- 이귀혜·강남준·이종영 (2008). 탄핵 시기 노무현 대통령의 수사학 연구: 컴퓨터 언어 분석 기법을 중심으로. 『한국언론학보』, 52권 5호, 25~55.
- 이진우 (2011). '위키리크스': 민주주의는 얼마나 많은 비밀을 감당할 수 있는가? 『철학과 현실』, 88-봄호, 121~131.
- 이태훈 (2010). 위키리크스사태는 디지털의 저주. 『주간조선』, 2010. 12. 13. 일자 기사.
- 윤영관 (2009). 오바마 행정부의 외교전략과 한반도. 『외교』, 89호.
- 조동원 (2011). 위키유출Wikileaks이 '폭로'한 대중간첩시대의 정보정치. 『황해문화』, 70호, 291~309.
- 홍성률 (2004). 1970년대 전반 북미관계: 남북대화, 미중관계 개선과의 관련 하에서. 『國際政治論叢』, 44집 2호, 29~54.
- Ambrose, Stephen E. (1993). *Rise to Globalism: American Foreign Policy Since 1938* (7th Ed.). 권만학 역(1997). 『국제 질서와 세계주의』. 서울: 을유문화사.
- Bantum, E. O. C. & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional

- expression in cancer narratives. *Psychological Assessment*, 21(1). 79~88.
- Bartels, Larry M. (2008). *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton: Princeton University Press.
- Benkler, Yochai (2011). A Free Irresponsible Press: Wikileaks and the Battle over the Soul of the Networked Fourth Estate. *Harvard Civil Rights-Civil Liberties Law Review*, 46, 311~397.
- Bennett, W. Lance (1990). Toward a Theory of Press-state Relations. *Journal of Communication*, 40(2), 103v125.
- Blei, David M., Ng, Andrew Y., & Jordan, Michael I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993~1022.
- Chandrasekaran, Pradeep K. K. (2012). *Measuring Interestingness of Documents Using Variability*. Unpublished Master Thesis. School of Computing. Queen's University. Kingston, Ontario, Canada.
- Cummings, Bruce (1981). *The Origins of the Korean War*. 김자동 역 (2001). 『한국전쟁의 기원』. 서울: 일월서각.
- Downs, Anthony (1957). *An Economic Theory of Democracy*. New York: Harper.
- Fenster, Mark (2011). *Disclosure's Effects: WikiLeaks and Transparency*. Unpublished Paper. Available at: http://works.bepress.com/mark_fenster/10
- Gelman, Andrew, Carlin, John B., Stern, Hal S., & Rubin, Donald B. (2004). *Bayesian Data Analysis*. New York: Chapman & Hall/CRC
- George, Alexander (1980). *Presidential Decisionmaking in Foreign Policy: The Effective Use of Information and Advice*. Boulder: Westview.
- Gilligan, James (2011). *Why Some Politicians Are More Dangerous Than Others*. 이희재 역(2011). 『왜 어떤 정치인은 다른 정치인보다 해로운가』. 교양인.
- Gitlin, Todd (1980). *The Whole World is Watching: Mass Media in the Making & Unmaking of the New Left*. Berkeley, CA: University of California Press.
- Goldstone, Jack A. (2011). Understanding the Revolutions of 2011. *Foreign Affairs*.
- Hallin, Daniel C. (1986). *The "Uncensored War": The Media and Vietnam*. New York: Oxford University Press.

- Hart, Roderick P. (1984). *Verbal Style and the Presidency: A Computer-based Analysis*. Orlando, FA: Academic Press.
- Hart, Roderick P. (2000). *DICTION 5.0: The Text-analysis program*. Thousands Oaks, CA: Scolari/Sage.
- Heisbourg, Francois (2011). Leaks and Lessons. *Survival: Global Politics and Strategy*. 53(1), 207~215.
- Keltner, Dacher & Lerner, Jennifer S. (2010). Emotion. In Fiske, Susan, Gilbert, Daniel T., & Lindzey, Gardner (Eds.), *Handbook of Social Psychology* (5th Ed., pp. 317~352). Hoboken, NJ: John Wiley & Sons, Inc.
- Keohane, Robert O. & Nye, Joseph S. (1977). *Power and Interdependence: World Politics in Transition*. Boston: Little Brown.
- Krippendorff, Klaus (2004). *Content Analysis: An Introduction to its Methodology*. Thousands Oaks, CA: SAGE Publications.
- Kim, John (2011). 『ウィキリークスからフェイスブック革命まで: 逆パノプティコン社会の到来』. 한석주·이단아 역 (2011). 『공개와 연대, 위키리크스와 페이스북의 정치학』. 서울: 커뮤니케이션 북스.
- Koh, Harold Hongju (2010. 11. 27). Why He Matters. *The Washington Post*. Available at: http://www.washingtonpost.com/politics/harold-hongju-koh/gIQAflj99O_topic.html
- Mayer, Rudolf & Rauber, Andreas (2011). On Wires and Cables: Content Analysis of Wikileaks Using Self-Organising Maps. In Laaksonen, Jorma & Honkela, Timo (Eds.) *Advances in Self-advanced Maps* (pp. 238~246). Berlin: Springer.
- McCraw, David J. (1994). New Zealan's Foreign Policy Under National and Labour Governments: Variations on the "Small State" Theme? *Pacific Affairs*, 67(1), 7~25.
- McElroy, Robert W. (1992). *Morality and American Foreign Policy: The Role of Ethics in International Affairs*. Princeton, N. J.: Princeton University Press.
- Morgenthau, H. J. (1948). *Politics among nations: The struggle for power and peace*. 이호재 역 (1987). 『現代 國際 政治論: 世界平和의 權力理論의 接近』. 서울: 법문사.
- Nye, Joseph S. (2006). *Soft power: The means to success in World Politics*. New York : Public Affairs.

- Opper, Melissa Hannah (2011). Wikileaks: Balancing First Amendment Rights with National Security. *Loyola of Los Angeles Entertainment Law Review*, 31(3), 237~267.
- Osgood, Charles E., Suci, George, & Tannenbaum, Percy. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Overy, R. J. (2004). *The Dictators: Hitler's Germany and Stalin's Russia*. Norton & Company.
- Page, Mark & Spence, J. E. (2011). Open Secrets Questionably Arrived At: The Impact of Wikileaks on Diplomacy. *Defense Studies*, 11(2), 234~243.
- Pang, Bo & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1~2), 1~135.
- Pennebaker, J. W. (1997). Writing about Emotional Experiences as a Therapeutic Process. *Psychological Science*, 8(3), 162~166.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R.J. (2007). The development and psychometric properties of LIWC2007. [Software manual]. Austin, TX.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological Aspects of Natural Language Use: Our Words, Our Lives. *Annual Review of Psychology*, 54(1), 547~577.
- Rachman, Gideon (2010. 12. 13). American Should Give Assange a Medal. *Financial Times*. Available at: <http://www.ft.com/cms/s/0/61f8fab0-06f3-11e0-8c29-00144feabdc0.html#axzz1qBUrRG2B>
- Riker, William H. (1982). The Two-Party System and Duverger's Law: An Essay on the History of Political Science. *The American Political Science Review*, 76(4), 753~766.
- Shane, Scott & Lehren Andrew (2000. 11. 29). *Leaked Cables Offers Raw Look at U.S. Diplomacy*. Available at: <http://www.nytimes.com/2010/11/29/world/29cables.html?pagewanted=all>
- Siebert, Frederick Seaton, Peterson, Theodore, & Schramm, Wilbur Lang (1956). *Four Theories of the Press: The Authoritarian, Libertarian, Social Responsibility, and Soviet Communist Concepts of What the Press Should Be and Do*. Chicago, IL, University of Illinois Press.
- Steyvers, Mark & Griffiths, Tom (2007). Probabilistic Topic Models. In

- Landauer, Thomas, K. McNamara, Danielle S., Dennis, Simon, & Kintsch, Walter (Eds.), *Handbook of Latent Semantic Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Susman, Louis (2010. 11. 28). U.S. *Ambassador Louis Susman on Potential Wikileaks Disclosures*. Available at:
<http://london.usembassy.gov/amb-speeches/susman023.html>
- Thayer, C. A. (1994). Sino-Vietnamese Relations: The Interplay of Ideology and National Interest. *Asian Survey*, 34(6), 513~528.
- Thompson, M., Zanna, M. P., & Griffin, D.W. (1995). Let's Not Be Indifferent about (Attitudinal) Ambivalence. In R. E. Petty, & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 361~386). Mahwah, J: Lawrence Erlbaum Associates.
- van Atteveldt, W., Kleinnijenhuis, J., & Ruigrok, N. (2008). Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles. *Political Analysis*, 16(4), 428~446.
- Waltz, K. N. (2000). Structural realism after the Cold War. *International Security*, 25(1), 5~41.
- Wei, William W. S. (1994). *Times Series Analysis: Univariate and Multivariate Methods*. Ney York: Addison-Wesley Publishing Company, Inc.
- Zaller, John (1992). *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

최초 투고일 2013년 12월 30일
 논문 수정일 2014년 02월 03일
 게재 확정일 2014년 02월 28일

A b s t r a c t

Computerized Text Analysis of Diplomatic Cables of US Embassy in South Korea

On the Effect of Political Changes in Country of Sojourn and Home Country on Topic Occurrences and Sentiment Expression

Young Min Baek

Assistant Professor, Department of Communication, Yonsei University

Moonho Choi

Ph. D Student, Department of Communication, Seoul National University

Ji-Yeon Jang

Ph. D Student, Department of Computer Science, Korea Advanced Institute of Science and Technology

This study conducts computerized text analysis of diplomatic cables released by American Embassy in South Korea, which is publicized on Wikileaks. Specifically, we analyzed total 1080 diplomatic cables, classified 'secret' or 'confidential,' ranging from January 1st 2012 to February 25th 2010. Interestingly, the period of publicized diplomatic cables on Wikileaks covered the political transition in both South Korea (country of sojourn) and the United States (home country). After conceptualizing the Korean transition as 'independent - following' and the US transition as 'unilateral - interdependence' we divided the whole period into three terms: (1) independent - unilateral, (2) following - unilateral, and (3) following - interdependence. We extracted seven topics in diplomatic cables, which are (1) national agenda in South Korea, (2) North Korea issue, (3) US forces in South Korea, (4) South-North economic cooperation, (5) Korean decision requesting troops, (6) East Asian issues, and (7) Korea-US trade relations. Also we quantified three types of sentiment expression: (1) positive, (2) negative, and (3) ambivalent sentiment using LIWC[®]2007 program. Two findings should

be noted. First, computerized text analysis warrants considerable level of validity of its results, at least topics and sentiment of a document. Second, change in topics and sentiment of diplomatic cables is more sensitive to political transition in the US than to that in South Korea.

KEYWORDS Wikileaks • Computerized Text Analysis • Latent Dirichlet Allocation • Principal Component Analysis • Sentiment Analysis • Time Series Analysis • Diplomatic Cables • Diplomatic Communication