

## 텍스트마이닝을 이용한 한국 대통령의 해양관에 관한 연구

김성국\* · 이태휘\*\*

### A Study on Text Mining Analysis of Presidential Maritime Concept in KOREA

Kim, Sung-Kuk · Lee, Tae-Hwee

#### Abstract

In the presidential political system, the word of the president has great influence on the formation of national policy and the decision-making process. Policy priorities are determined according to the president's ideology and core values, and various policies are established and executed according to the priorities. Therefore, this paper analyzes the contents of the president's speech.

Since the president's speech is a semantic datum, in order to analyze unstructured text, big data analysis is conducted through the methods of machine learning and deep learning. In this study, the president's speech at the "National Sea Day" commemoration was obtained 1996 onwards and analyzed using topic modeling.

As a result of the analysis, all the presidents' speeches were delivered with a view of the ocean that was consistent with the direction of their administration. It was confirmed that the ocean-industry-resource topics, which are the intrinsic values of the ocean, were not damaged and consistently emphasized by all presidents.

*Key words: Text Mining, Speech, Topic Model, Content Analysis*

▷ 논문접수: 2020. 04. 08.   ▷ 심사완료: 2020. 06. 17.   ▷ 게재확정: 2020. 09. 23.

\* 경남과학기술대학교 유통학과 강사, 제1저자, altasea@naver.com

\*\* 경남과학기술대학교 상경대학 조교수, 교신저자, taylor@gntech.ac.kr

## I. 서론

대통령 중심제(presidential system)의 정치체제에 서는 대통령의 언어가 국가정책의 형성과 의사결정 과정에 지대한 영향을 미치게 되어 있다(Linz, 1985). 대통령의 이념과 중심가치가 무엇인가에 따라 정책우선순위가 결정되고 그 우선순위에 따라 다양한 정책이 수립되고 집행된다.

그래서 대통령의 문체 해석력과 의식지향을 파악하고자 하는 노력이 여러 연구자들에 의해 진행되고 있다. 이 가운데 대통령의 연설문을 분석함으로써 대통령의 가치관, 사고, 사상을 연구하는 연구가 우리나라에서도 진행되고 있다(김동환, 2000; 이창길, 2010; 박준형·유호선·김태영·한희정·오효정, 2017; 김성국·이태휘, 2019).

대통령의 연설문을 분석하기 위해서는 내용분석(content analysis) 방법을 이용하여 각종 자료의 내용을 구체적이고 세밀하게 조사하여야 하지만 정성적 자료를 정량화 시키는데 어려움이 있었다. 전통적인 사회과학적 연구에서는 이론을 중요시하며 이론을 증명하기 위해서는 인과관계 추론(causal inference)이 중요하였다. 이론, 방법, 자료가 과학적으로 마련되어야 하고 관찰된 자료로부터 원인과 결과의 관계라는 결론을 도출하는 과학적 연구에서는 상관관계를 보여주는 내용분석은 주류 연구방법에서는 벗어나 있었다.

하지만 빅데이터(big data)의 시대가 도래하여 슈퍼컴퓨터 등의 정보기술의 발전으로 전통적 사회과학연구 방법의 한계를 넘어서는 방법론이 도입되고 있다. 특히 오늘날 빅데이터의 80% 이상이 텍스트 등의 비정형 데이터로 구성되어 있다(Economist, 2015). 텍스트는 컴퓨터로 처리될 수 있는 형태로 저장된 글, 곧 언어 자료이다. 언어자료는 비정형

데이터이기 때문에 텍스트 자료의 분석론이 중요해지고 있다.

텍스트 자료 분석은 대용량의 데이터베이스로부터 의미 있는 규칙과 패턴을 발견하는 데이터마이닝(data mining)의 일종이다(Hearst, 1999). 비정형, 비구조화 된 언어 자료를 사용하여 의미 있는 구조와 상관관계를 밝히는 연구방법으로서 토픽모델(topic model)이 이용되고 있다. 기계학습(machine learning) 기법의 개발은 문장속에 나타난 단어를 분석하여 핵심 의미를 파악하는 분석을 수월하게 지원하고 있다.

한편 3면이 바다로 둘러쌓여 있으며 바다를 통해 국제무역을 통해 국부를 창출하는 우리나라에 있어서 해양 혹은 바다는 국가적으로 중요한 의미가 있다. 세계 선진국들이 바다와 관련한 기념일을 갖고 있는데 우리나라 역시 바다의 날을 제정하여 5월 31일을 매년 기념하고 있다. 국가적인 기념일이기 때문에 기념식에는 대통령이 참석하여 기념식 연설을 하는 국가행사 개최되고 있다. 이러한 국가 기념일에 사용되는 대통령의 기념식 연설문은 대통령의 해양관을 확인할 수 있는 중요한 자료가 된다.

이에 본 연구에서는 바다의 날에 대통령이 사용한 연설문을 중요한 자료로 판단하고 연설문에서 의미있는 구조를 파악하고자 한다. 이를 위하여 1996년부터 24년동안 사용된 바다의 날 기념식의 대통령 연설문을 확보하고 텍스트마이닝을 시도하였다.

본 연구의 제 I 장에서는 연구배경과 연구목적을 제기하고 제 II 장에서 대통령의 연설문을 분석하는 선행연구와 방법론을 검토하였다. 텍스트마이닝 연구방법을 이용하여 제 III 장에서 우리나라의 역대 대통령의 연설문을 분석하고 제 IV 장에서 해양관에 대한 의미를 확인하였으며 제 V 장에서 결론을 내린다.

## II. 이론적 배경

### 1. 연설문의 의의

대부분의 정부의제 형성과정은 먼저 이슈가 제안되고 그 해결책이 구체화되고 그 이슈에 대한 지지가 확산되어 성공하면 이슈가 정부의제로 편입되게 된다. Cobb, Ross and Ross(1976)의 의제형성 모형에 따르면 공중의제가 공식의제가 되는 과정의 특징에 모형은 외부주도형(outside initiative model), 동원형(mobilization model), 내부주도형(inside access model)으로 분류된다.

여기에서 동원형은 정부 내의 정책결정자(통치권자나 최고관리자)들에 의하여 의제화가 주도되는 경우 주로 정치지도자들의 지시에 의하여 사회문제가 바로 정부의제로 채택되고, 일반대 중의 지지를 얻기 위해 정부의 의도적인 노력에 의해 공중의제로 확산된다.

또한 May(1991)는 이미 대중의 지지가 높은 정책문제에 대하여 정부가 그 과정을 주도하여 해결을 시도하는 유형으로 공고화 모형(consolidation model)을 추가 하였다.

의제설정과정의 유형을 정리하면 <표 1>과 같은데, 정부가 대중의 지지를 결합하고자 하는 이유는 정책이 결정된 이후 집행하는 것이 유리하기 때문이다.

표 1. 의제설정과정의 유형

		대중의 관여도	
		높음	낮음
의제설정의 주도자	민간	외부주도	내부주도
	정부	공고화	동원

자료: Howlett, and Ramesh(2003), p.141.

특히 대통령 연설문은 국가비전, 국정철학, 정책기조, 중심가치 등을 국민들에게 전달하는 수단이 된다(박준형·유호선·김태영·한희정·오효정, 2017)

우리나라와 같은 대통령 중심제 정치체제를 갖고 있는 나라에서는 한 국가정책의 형성과 정책 의사결정 과정에 미치는 영향은 매우 지대하다. 대통령 중심제 하에서는 지도자인의 이념적 경향과 리더십 스타일 및 정책의제에 대한 인식과 가치에 따라 정책방향은 여러 가지 방향으로 표출될 수밖에 없다(김혁, 2014).

중심 가치를 내포하는 다양한 지표들 중에서도 대통령 연설문은 대통령이 국가비전, 국정철학, 정책기조 등을 국민들에게 전달하기 위해 자주 활용하는 수단으로서, 해당 정부의 국정목표 및 국정철학을 내포하고 있다(이창길, 2010).

박준형·유호선·김태영·한희정·오효정(2017)은 대통령 연설기록물의 중심가치 연구를 분석한 바 있다. 분석방법으로는 워드클라우드(word cloud), 단어 빈도분석, 단어 간 연계구조 분석을 실시하였으며 중심가치에 대한 명확한 정의는 없이 자주 등장하는 단어들을 대통령과 정부의 중심가치로 이해하였다.

이창길(2010)은 노무현 대통령의 참여정부와 이명박 대통령의 실용정부에서 발표한 연설문을 대상으로 언어 네트워크 분석(semantic network analysis)을 수행하고 정부 초기의 가치지향과 정책우선순위를 파악하였다.

김동환(2000)은 imf 금융위기의 이후 김대중 대통령의 사고를 인과지도(casual map)로 분석하기 위해 대통령 취임 이후 인터뷰, 연설문, 국무회의, 간담회 발언을 토대로 분석하였다.

김성국·이태휘(2019)은 국가기록원 대통령기록관에서 제공하는 주요한 물류관련 기념식에서 발표한 연설문 14개 연설문을 발췌하여 분석하였다. 분



념일 등에 관한 규정(대통령령 제27619호, 2016.11.30.시행) 제2조제1항에 의해 주관부처는 해양수산부이며 “바다 관련 산업의 중요성과 의의를 높이고 국민의 해양사상을 고취하며, 관계종사원들의 노고를 위로하는 행사” 를 한다. 1996년 제1회 바다의 날 기념식이 부산 신선대부두에서 김영삼 대통령이 참가하면서 대통령 행사로 개최되었다.

2019년 현재 22번의 바다의 날 기념식이 <표 3> 과 같이 개최되었다. 2009년 14회와 2014년 19회 기념식은 각각 노무현대통령 국민장과 세월호 사고로 행사식은 개최되지 않았지만 정부행사로서는 가장 큰 해양 기념식이다.

표 3. 바다의 날 기념식 현황

연도	회수	정부수반	기념식 행사장
1996	1회	김영삼	부산 신선대부두
1997	2회	김영삼	광양 제철부두
1998	3회	김대중	부산 감만부두
1999	4회	김대중	마산 제5부두
2000	5회	김대중	서울 세종회관
2001	6회	김대중	여수 오동도
2002	7회	김대중	서울 ASEM회의장
2003	8회	노무현	인천 제5부두
2004	9회	노무현	보령 대천해수욕장
2005	10회	노무현	울산 해양공원
2006	11회	노무현	서울 코엑스
2007	12회	노무현	포항 영일만신항
2008	13회	이명박	여주시 오동도
2009	14회	이명박	노무현대통령 국민장으로 취소
2010	15회	이명박	서울 COEX
2011	16회	이명박	강원 화진포해수욕장
2012	17회	이명박	여수 박람회장
2013	18회	박근혜	태안 만리포해수욕장
2014	19회	박근혜	세월호 사고로 취소
2015	20회	박근혜	부산 국립해양박물관
2016	21회	박근혜	순천 순천만국가정원
2017	22회	문재인	군산 새만금광장
2018	23회	문재인	인천 내항8부두
2019	24회	문재인	울산 미포조선

자료: 저자 작성.

대통령이 바다의 날 기념식에 참석하여 발표하는 연설문은 평소 가지고 있는 신념과 이념, 가치관 등이 압축된 해양관을 나타내고 있을 것으로 짐작할 수 있다. 대통령의 연설문을 분석한 선행 연구들의 대부분이 국정 전반에 걸친 특정 대통령의 가치관과 인식을 분석한 것이 대부분이었다.

바다라는 정책적 공간에서 연속된 정부의 인식을 살펴봄으로써 해양국가인 우리나라에서 대통령의 해양관을 살펴보는 것은 의미가 있다.

### III. 토픽모델 분석

#### 1. 텍스트마이닝 연구방법

내용분석(content analysis)은 연설문, 서한, 기록 문서, 신문 등과 같은 각종 자료의 내용을 구체적이고 세밀하게 조사하고 내용 속에 함축된 메시지를 분류하거나 빈도수를 측정하여 내용 속에 담겨진 의미나 특성을 파악해 내는 방법이다. 하지만 전통적인 내용분석 방법은 외적 타당성의 제한, 고비용의 연구 환경, 개념적 모호성, 연구자에 의한 분석항목의 과도한 의존성 등이 한계점으로 지적되고 있다(Danowski, 1993).

전통적인 ‘과학적 방법’은 관찰된 자료로부터 두 변수 사이의 관계가 원인과 결과의 관계를 규명하기 위하여 이론에 근거한 가설과 모형을 제시하였다(Box, 1987). 결론은 엄밀한 사고를 통해 도출해내는 과정으로서의 인과관계 추론(causal inference)을 중심에 두고 있다(Hume, 1978).

특히 자료에서 경험적으로 찾는 상관관계는 인과관계를 규정하는데 있어서 필요조건일 뿐이라는 견해였다(Song, Kim and Bhattacharjee, 2014), 따라서 전통적인 사회과학적 연구는 어떻게 원인이 결과를 만들어내는지를 설명하고, 그 기제를 제시하는

이론적 논의가 있어야 한다(Marini and Singer, 1988).

과거의 물질적, 기술적 제약들이 조건지었던 전통적 사회과학적 연구방법의 한계를 슈퍼컴퓨터, 기계학습, 인공지능(artificial intelligence)이 극복하고 있다. 이러한 기술은 지금까지 상상하지 못한 정도로 정교하게 추적하고 측정할 수 있게 되었기 때문에 전통적 사회과학적 연구의 사고에서 벗어날 수 있다(한신갑, 2015).

Halevy, Norvig and Pereira(2009)는 빅데이터 분석과 같은 연구방법으로 인해서 양적, 질적으로 근본적으로 달라진 자료가 극단적 형태의 귀납적 경험주의를 가능하게 만들었다고 주장하면서 이제 상관관계만으로도 충분한 시대가 되었다고 주장한다.

데이터가 충분히 주어진다면 숫자가 스스로 자기 얘기를 할 수 있게 된다(Anderson, 2008). 상관관계가 인과관계를 대체하게 되고, 일관성을 갖춘 모형이나 통합된 이론 등의 체계적 설명이 없이도 과학은 발전해나갈 수 있다.

따라서 본 연구에서 시도하는 텍스트 마이닝은 전통적인 사회과학연구 방법적 접근이 아니라고 하더라도 선행연구의 관계성을 확보하고 분석하고자 한다. Lazer, Pentland, Adamic, Aral, Barabási, Brewer, Christakis, Contractor, Fowler, Gutmann, Jebara, King, Macy, Roy and Van Alstyne(2009)의 주장과 같이 그동안 모르고 있었던 방법으로 인간 세계를 관찰 할 수 있다는 기대가 널리 퍼져 있는 만큼 새로운 시대의 분석방법으로 연구를 진행하고자 한다(<그림 2> 참고).

그림 2. 자료의 유형과 분석

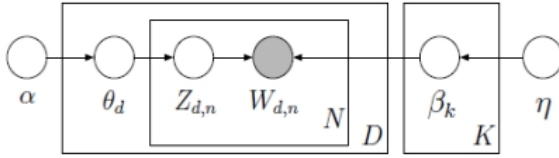
Style of research	Source of evidence	Type of data	Type of analysis
Survey research (조사 연구)	Questionnaires, interviews (설문, 인터뷰)	Attribute (속성)	Variable analysis (변수분석)
Ethnography research (민속학 연구)	Observations (관찰)	Ideational (관념적)	Typological analysis (분류분석)
Documentary research (문헌 연구)	Texts (텍스트)	Relational (관계적)	Network analysis (네트워크분석)

자료: Scott(2000), p.3.

빅데이터 연구에서 대량의 문서에서 주제를 추출하는 방법론으로 토픽모델링을 자주 이용한다. 토픽 모델링이란 기계 학습 및 자연어 처리 분야에서 토픽(topic, 주제어)이라는 문서 집합의 추상적인 주제를 발견하기 위한 통계적 모델로서 텍스트 본문의 숨겨진 의미 구조를 발견하기 위해 사용되는 텍스트 마이닝(text mining) 기법이다(Steyvers and Griffiths, 2007).

토픽모델링의 대표적인 기법인 LDA(latent dirichlet allocation) 분석은 <그림 3>과 같이 문서 안에 존재하는 주제들을 찾아내는 기법으로, 문서 집합의 주제와 각 문서별 주제 비율, 각 주제에 포함될 단어들을 도출해낼 수 있다.

그림 3. 토픽모델(LDA) 구조



- $K$  : 토픽의 수
  - $\alpha$  : 문서별 토픽  $k$ 의 Dirichlet prior weight,  $\theta$  값을 결정하는 파라미터
  - $\eta$  : 토픽별 단어  $w$ 의 Dirichlet prior weight,  $\beta$  값을 결정하는 파라미터
  - $\theta_d$  : 문서별 토픽의 비율
  - $\beta_k$  : 문서별 단어  $w$ 의 생성확률
  - $Z_{d,n}$  : 문서  $d$ 의  $n$ 번째 단어의 토픽(index)
  - $W_{d,n}$  : 문서  $d$ 의  $n$ 번째 단어(문서에 관측되는 변수, index)
- 자료: Blei, Ng and Jordan(2003).

LDA는 다수의 문서에서 관찰되는 변수를 통해서 보이지 않는 변수를 추론하여 문서에 숨겨져 있는 주제를 발견하는 절차적 확률 분포로서 변수는 단어( $w_{d,n}$ )로 관측된 데이터이다.  $\alpha$ ,  $\beta$ 는 단어를 추출하기 위해 사용된 모수이고  $\theta$ ,  $Z$ 는 직접 관찰할 수 없거나 측정되지 않은 잠재변수(latent variable)이다.<sup>1)</sup> LDA에서 토픽별 단어 생성확률인  $\beta$ 의 Dirichlet 분포가 결정되고, 각 단어의 토픽을 나타내는 값인  $Z$ 와 토픽별 단어 비율인  $\beta$  값에 따라 단어  $W$ 가 결정된다.<sup>2)</sup>

## 2. 토픽 모델 분석

1) 초매개변수(hyperparameter)인  $\alpha$ ,  $\eta$ 와 단어를 추출하기 위해 사용되는 숨겨진 매개변수(hidden parameter)인  $\beta$ 가 있다. 머신러닝(machine learning)에서 hyperparameter는 신경망 학습을 통해서 튜닝 또는 최적화 해야하는 주변수가 아니라 학습 진도율이나 일반화 변수처럼, 사람들이 선택적 지식으로 설정을 하거나 또는 외부 모델 메커니즘을 통해 자동으로 설정이 되는 변수를 말한다. Aggarwal(2018), p.125.

2) LDA에서  $Z_d, \eta | \theta$  가 다항분포를 따르기 때문에 Dirichlet 분포를 사용하게 되면 사후확률의 분포는 다항분포가 된다. 보다 상세한 LDA 모델의 도출 함수식은 Blei, Ng and Jordan(2003)의 제안모델에서 확인할 수 있다.

본 연구에서의 연구목표가 역대 대통령의 해양관 인식에 관한 연구이기 때문에 <표 4>와 같은 기초 자료를 토대로 토픽 분석에 사용하였다.

토픽 분석에 사용한 도구는 상용 프로그램인 Cyram(2015)의 NetMiner 4(Ver.4.4)를 사용하였다. NetMiner는 소셜 네트워크 분석용 상용 소프트웨어로서 한국어 처리가 가능하다. 최근 자연어 처리와 딥러닝(deep learning)에 적합한 파이썬(Python) 기반 스크립트와 플러그인을 제공하고 있다.

본 연구에서 분석에서 사용하는 대통령의 바다의 날 연설문은 2019년 기준으로 11개를 확보 할 수 있었다. 대통령이 주빈으로 참석하지 않았던 기념식에서는 총리 기념사 7개, 장관 기념사 1개를 확보하여 사용하였다.

한편 국가기록원(www.archives.go.kr)에서 대통령의 바다의 날 기념사를 확보하지 못한 시기에는 같은 연도에 대통령이 바다 관련 행사에 참석하여 사용한 기념사를 반영하였다. 구체적으로 해군사관학교 졸업식 2회, 및 해양경찰의 날, 부산 센텀시티 기공식, 남극과학기지 준공식 연설문 등 5개를 대통령의 연설문으로 반영하였다.

표 4. 바다의 날 연설문 기초자료

연도	회수	연설문	비고
1996	1회	대통령	
1997	2회	대통령	
1998	3회	대통령	
1999	4회	대통령	해군사관학교 졸업식 연설 (1999.3.12.)로 대체
2000	5회	대통령	부산 센텀시티 기공식 연설 (200.11.2)로 대체
2001	6회	대통령	
2002	7회	대통령	해군사관학교 졸업식 연설 (2002.3.14.)로 대체
2003	8회	대통령	
2004	9회	대통령	
2005	10회	대통령	
2006	11회	총리	
2007	12회	총리	
2008	13회	대통령	
2009	14회	장관	
2010	15회	총리	
2011	16회	대통령	해양경찰의 날 기념식 연설 (2011.9.16.)로 대체
2012	17회	총리	
2013	18회	총리	
2014	19회	대통령	남극과학기지 준공식 연설 (2014.2.10.)로 대체
2015	20회	대통령	
2016	21회	대통령	
2017	22회	대통령	
2018	23회	총리	
2019	24회	총리	

연설문 합계: 24(대통령: 16, 총리: 7, 장관: 1)

토픽모델링을 이용한 분석결과는 <표 5>와 같이 연도별로 정리하였다. 토픽의 개수는 4개로 제한하여 설정하였다. 대부분의 연구논문에서 키워드를 4~5개를 압축하여 제시하는 것이 관례화 되어 있기 때문에 인간의 표현을 압축하는 토픽으로 4개로 제한하는 것은 큰 무리가 없다.

실무적으로는 너무 적은 토픽은 의미가 중복되어 해석하는 것이 곤란하고 토픽이 너무 세분화 될 경우에는 사용 빈도가 줄어든다는 경험치가 있어서 분석도구에서도 설정값(default)이 4개이다.

연도별 분석을 보면 1996년에는 제1회 바다의 날이 지정되고 1996년 8월 8일에 해양수산부가 대통령령 제15135호에 의해 발족되기에 해양부라는 단어가 등장하였고 가장 많은 빈도를 보였다.

표 5. 연도별 대통령의 해양관 토픽

연도	토픽	Keyword					빈도
		1st	2nd	3rd	4th	5th	
1996	Topic1	국가	세계	건설	일류	중심	10
	Topic2	해양	산업	환경	관계자	해양부	14
	Topic3	해양	시대	세기	나라	세계	9
	Topic4	바다	민족	인류	시작	국민	15
1997	Topic1	바다	국가	경영	국민	동아시아	8
	Topic2	바다	세계	미래	나라	시대	8
	Topic3	해양	강국	건설	바다	조국	6
	Topic4	수산	해양	가족	자리	모두	7
1998	Topic1	바다	시대	해양	개혁	세기	13
	Topic2	경제	극복	난국	발전	부산	5
	Topic3	해양	추진	수산	나라	바다	14
	Topic4	세계	한국	물류	중심지	물자	10
1999	Topic1	해군	바다	장교	장병	졸업	19
	Topic2	조국	바다	졸업	장보고	동아시아	6
	Topic3	안보	태세	대비	국가	바다	8
	Topic4	북한	안보	위협	대북	도발	17
2000	Topic1	부산	시민	시티	센텀	착공	12
	Topic2	세계	부산	경제	나라	정부	15
	Topic3	아시아	정보	유통	연결	산업	12
	Topic4	남북	관계	정상	협력	회담	26
2001	Topic1	바다	세계	박람회	해양	여수	27
	Topic2	해양	세기	태평양	산업	자랑	19
	Topic3	어업	확충	거점	물류	관리	8
	Topic4	국민	바다	존경	나라	여수	21
2002	Topic1	국민	평화	남북	북한	이후	22
	Topic2	세계	월드컵	안전	한국	경제	18
	Topic3	해군	바다	시대	조국	해군력	20
	Topic4	친애	졸업	자리	협력	협력	33
2003	Topic1	세계	바다	동안	시대	동북아	17
	Topic2	자원	미래	에너지	환경	개발	7



	Topic3	동북아	중심	물류	경제	항해	11
	Topic4	해양	국민	역사	존경	등대	19
2004	Topic1	바다	나라	해양	국가	가능	9
	Topic2	감사	인사	지원	이번	행사	2
	Topic3	어촌	수산업	자율	사례	활로	2
	Topic4	바다	해양	수산	나라	수산업	4
2005	Topic1	해양	기술	개발	자원	수산	4
	Topic2	해양	세계	감사	수산	발전	4
	Topic3	바다	축하	도전	발전	감사	6
	Topic4	바다	세계	도전	감사	축하	4
2006	Topic1	해양	수산	국민	존경	강국	10
	Topic2	정부	어려움	정성	어업	여건	3
	Topic3	바다	미래	희망	올해	축하	9
	Topic4	세계	수산물	양력	생산량	조선	4
2007	Topic1	해양	바다	개발	물류	도약	5
	Topic2	해양	바다	포항	발전	개발	2
	Topic3	세계	바다	해양	도약	나라	13
	Topic4	바다	세계	발전	포항	나라	7
2008	Topic1	해양	바다	국가	확보	보호	10
	Topic2	감사	모두	바다	기원	산인	3
	Topic3	해양	세계	노력	자원	개발	6
	Topic4	선진	부국	건설	정부	박람회	2
2009	Topic1	해양	화성	레저	최영근	시장	8
	Topic2	바다	기념식	자리	세계	지역	16
	Topic3	세계	내외	귀빈	녹색	성장	17
	Topic4	해양	레저	감사	가능	지속	67
2010	Topic1	국민	물류	조성	환경	존경	5
	Topic2	세계	여수	성공	계기	역사	5
	Topic3	해양	성장	세계	분야	자원	7
	Topic4	바다	개척	해양	감사	창출	11
2011	Topic1	해양경찰	국민	축하	불법	조업	16
	Topic2	해양	관광	산업	구조	문화	9
	Topic3	바다	경제	세계	한반도	아시아	19
	Topic4	해양경찰	정부	선진	사랑	대원	16
2012	Topic1	세계	바다	박람회	이번	여수	9
	Topic2	해양	바다	감사	자리	환경	8
	Topic3	자원	바다	환경	금속	석유	3
	Topic4	해양	산업	과학	세계	물류	6
2013	Topic1	바다	국민	미래	해양수산	사람	5
	Topic2	해양	수산	과학	영토	산업	8
	Topic3	바다	태안	희망	자원	사고	9
	Topic4	산업	개발	북극권	항만	미래	4
2014	Topic1	자원	기회	과학	해양	간직	3
	Topic2	과학	기지	축하	준공	남극	4
	Topic3	나라	연구	개척	극지	영토	2
	Topic4	장보고	영토	개척	남극	연구	0

2015	Topic1	중심	해양수산	선도	클러스터	지원	1
	Topic2	국민	바다	휴식처	수산물	먹거리	1
	Topic3	해양	대한민국	강국	열정	세계	4
	Topic4	바다	국민	축하	메시지	영상	3
2016	Topic1	세계	바다	사례	모범	환경	1
	Topic2	바다	발전	성장	축하	미래	5
	Topic3	해양	발전	세계	경제	사례	3
	Topic4	바다	산업	해양	발전	축하	0
2017	Topic1	해양	수산	선박	정부	지원	23
	Topic2	국민	세계	나라	존경	세월	14
	Topic3	산업	해운	조선	정책	국가	14
	Topic4	바다	경제	안보	세만금	국가	35
2018	Topic1	바다	가치	인류	활용	국가	17
	Topic2	체계	공간	수출	가공	발전	7
	Topic3	해양	산업	해운	항만	회복	7
	Topic4	해양	수산	인천	바다	도전	21
2019	Topic1	플라스틱	바다거북	발견	벡속	국민	6
	Topic2	활용	지역	항만	지원	국민	9
	Topic3	바다	인류	도약	고갈	안전	15
	Topic4	해양	쓰레기	정부	플라스틱	안전	22

## IV. 대통령의 해양관

### 1. 대통령의 연설문 분석

대통령이 발표한 바다의 날 연설문을 토픽별로 분석하면 <표 6>과 같다. 김영삼 대통령은 바다-미래-민족, 해양-산업-환경, 해양-시대-세계, 바다-국가-국민의 토픽으로 언급하였다. 이 가운데 해양-시대-세계가 가장 높은 빈출을 보였는데 이것은 당시 “세계화 선언” 등의 국정철학과 연관되는 주제와 일맥상통한다.<sup>3)</sup>

김대중 대통령은 부산-시민-해양, 국민-남북-협력, 해양-세계-바다, 바다-해군-국가의 토픽으로 언급하였다. 바다의 날 기념사를 확보하지 못하여 해군사

3) 김영삼민주센터(<http://kyscd.org>)의 분석에 따르면 김영삼 대통령의 주요업적으로 문민시대 개막, 깨끗한 경제, 정치 개혁 제도화, 역사바로세우기, 세계화 추구를 들고 있다.

관학교 졸업식 및 부산 센텀시티 기공식의 자료를 사용하여 과대 추정되었을 것으로 생각하면 남북화해 시대의 메시지로써 국민-남북-협력의 주제가 비교적 많이 사용되고 있다.<sup>4)</sup>

노무현 대통령은 바다-미래-역사, 해양-수산-자원, 세계-해양-국민, 동북아-바다-물류라는 토픽으로 연설문이 구성되어 있다. 전임 대통령과 달리 “물류”라는 용어를 사용하였고 가장 빈출한 토픽으로 나타난다. 이것은 노무현 대통령이 2000년 8월 7일부터 2001년 3월 25일까지 해양수산부장관을 역임한 경험과 동북아물류중심국가의 국정주제와 일치하는 토픽이다.<sup>5)</sup>

이명박 대통령은 바다-세계-나라, 해양경찰-국민-감사, 바다-해양-국가, 해양-자원-산업의 토픽을 주로 이용하였는데 2011년 16회 바다의 날 기념사 대신에 해양경찰의 날 기념식 연설로 대체하여 해양경찰 토픽이 과대 추정된 것으로 보인다. 하지만 해양-자원-산업의 토픽을 가장 많이 나타냈는데 이것은 이명박 정부의 자원외교 정책과 일치한다.<sup>6)</sup>

박근혜 대통령은 바다-미래-성장, 해양-산업-수산, 과학-자원-기지, 바다-축하-태안의 토픽이 자주 등장하고 있다. 박근혜 대통령 재임시에 2014년에는 세월호 사고가 있어서 바다의 날 기념사 대신에 남극 과학기지 준공식 연설로 대체하여 과학 토픽이 과대 추정된 것으로 보인다. 해양-산업-수산의 전형적

인 토픽이 가장 많이 이용하였고 국정비전 이었던 “국민행복, 희망의 새 시대”의 단어인 “희망”, “국민”이 등장하였지만 국정목표였던 “창조경제”는 나타나지 않았다.<sup>7)</sup>

2019년 까지 문재인 대통령의 해양관은 바다-인류-활용, 산업-해운-해양, 바다-국민-플라스틱, 해양-수산-정부의 토픽을 형성하고 있다. 해양-수산-정부 토픽이 가장 많이 사용되고 있으며 다음으로 지구 온난화 문제 등으로 인하여 바다-인류-활용 토픽이 많이 사용되고 있다. 최근 문제가 되고 있는 해양 플라스틱에 관한 단어가 등장하는 등 환경보존과 관련된 토픽이 많이 나타나고 있다. 또한 대통령의 연설문 가운데 독특하게 나타난 “정부”라는 단어는 국정운영과제 첫 번째 목표가 “국민이 주인인 정부”를 반영하고 있다.<sup>8)</sup>

표 6. 대통령의 해양관 토픽

	토픽	Keyword					빈도
		1st	2nd	3rd	4th	5th	
김영삼	Topic1	바다	미래	민족	인류	후손	16
	Topic2	해양	산업	환경	관계자	행정	21
	Topic3	해양	시대	세계	세기	나라	22
	Topic4	바다	국가	국민	건설	일류	18
김대중	Topic1	부산	시민	해양	세계	친애	33
	Topic2	국민	남북	협력	안보	관계	71
	Topic3	해양	세계	바다	세기	시대	77
	Topic4	바다	해군	국가	세계	자리	144
노무현	Topic1	바다	미래	역사	발전	기회	35
	Topic2	해양	수산	자원	개발	관리	32
	Topic3	세계	해양	국민	강국	존경	28
	Topic4	동북아	바다	물류	중심	축하	47

4) 김대중평화센터(<http://www.kdipeace.com>)의 따르면 6.15 남북정상회담 등 평화정착에 기여한 것을 주요한 업적으로 들고 있다.

5) 노무현 대통령은 취임사에 동북아 경제중심국가론을 밝히는 등 동북아 공동체 건설에 강한 의지를 가지고 있었다(한겨레신문, 2003.2.25.).

6) 이명박 대통령은 2007년 서울외신기자클럽 간담회에서 자신의 외교·통일 정책 구상을 발표하면서 안정적인 자원 확보를 위한 '에너지 실크로드' 구축을 밝혔고 대통령 당선된 직후 '자원개발 지주회사'를 통한 대형화 방침으로 본격화되었다(최원기, 2013).

7) 박근혜 대통령의 국정비전은 “국민행복, 희망의 새시대”으로 정하고 첫 번째 국정목표로 “일자리 중심 창조경제”를 제시하여 창조경제에 대하여 강한 의욕을 보였다(차두원·유지연, 2013).

8) 문재인 대통령의 국정 목표는 ① 국민이 주인인 정부, ② 더불어 잘사는 경제, ③ 내 삶을 책임지는 국가, ④ 고르게 발전하는 지역, ⑤ 평화와 번영의 한반도이다(국정기획자문위원회, 2017).

이명박	Topic1	바다	세계	나라	경제	박람회	58
	Topic2	해양경찰	국민	감사	축하	귀빈	48
	Topic3	바다	해양	국가	선진	감사	22
	Topic4	해양	자원	산업	세계	개발	115
박근혜	Topic1	바다	미래	성장	국민	해양수산	9
	Topic2	해양	산업	수산	발전	경제	18
	Topic3	과학	자원	기지	남극	장보고	10
	Topic4	바다	축하	태안	희망	세계	16
문재인	Topic1	바다	인류	활용	경제	국가	48
	Topic2	산업	해운	해양	조선	선박	36
	Topic3	바다	국민	플라스틱	쓰레기	해양	30
	Topic4	해양	수산	정부	감사	수산업	76

표 7. 한국 대통령의 해양관 토픽

토픽	Keyword					빈도
	1st	2nd	3rd	4th	5th	
Topic1	국민	해양	바다	존경	수산	150
Topic2	세계	바다	해양	경제	물류	169
Topic3	바다	감사	축하	미래	나라	222
Topic4	해양	산업	자원	수산	세계	489

분석도구에서 제시하는 키워드의 토픽 동시 출현 확률(topic probability)과 토픽 연결정도(topic degree)는 <표 8>과 같이 제시하고 있다. 연설문에 사용된 단어는 총 1,637개 이며 가장 많이 나타난 단어는 바다(302개)이며 다음으로 해양(295개), 세계(152개), 국민(79개), 수산(69개)의 순서로 나타나고 있다.

2. 한국 대통령의 인식

대통령의 24개 연설문을 분석한 각 토픽별 키워드는 <그림 4>와 같은 워드클라우드로 나타낼 수 있다. 토픽에 등장하는 단어의 빈출빈도에 따라 큰 글씨로 나타나는데 해양, 바다, 수산, 국민, 감사, 시민, 존경 등의 키워드가 두드러져 있다.

그림 4. 대통령의 해양관 워드클라우드



대통령의 바다의 날 기념사 연설문을 모두 통합하여 평가하였을 경우에는 <표 7>과 같이 국민-해양-바다, 세계-바다-해양, 바다-감사-축하, 해양-산업-자원 토픽으로 형성되고 있다.

이 가운데 해양-산업-자원-수산-세계 토픽이 가장 많이 이용되고 있는데 이것은 우리나라 대통령들이 가장 많이 인식하고 있는 토픽으로 확인된다.

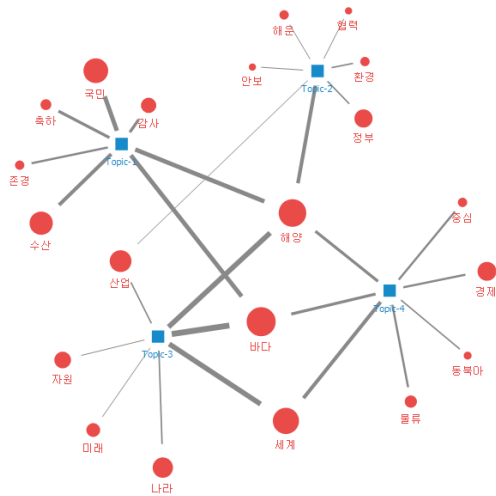
표 8. 토픽 키워드 구성

단어	빈도	동시출현 확률				Topic degree
		Topic1	Topic2	Topic3	Topic4	
해양	295	0	0	0.1	0	4
바다	302	0	0	0.1	0	3
세계	152	0	0	0	0	2
산업	66	0	0	0	0	2
국민	79	0	0	0	0	1
수산	69	0	0	0	0	1
나라	56	0	0	0	0	1
경제	55	0	0	0	0	1
정부	54	0	0	0	0	1
자원	50	0	0	0	0	1
감사	44	0	0	0	0	1
미래	42	0	0	0	0	1
물류	37	0	0	0	0	1
축하	35	0	0	0	0	1
존경	31	0	0	0	0	1
중심	31	0	0	0	0	1
해운	31	0	0	0	0	1
환경	31	0	0	0	0	1
동북아	28	0	0	0	0	1
안보	24	0	0	0	0	1
협력	24	0	0	0	0	1

한편 빈출된 토픽 구조가 어떻게 연결되어 있는지를 파악하기 위해서는 네트워크를 파악함으로써 가능하다. 각 토픽에 이용된 키워드를 네트워크로 나타내면 <그림 5>와 같이 나타난다.

연설문에서 빈도가 가장 많이 등장하고 있는 키워드는 “바다”이다. 이에 비하여 “해양”은 바다 보다는 빈도수가 낮았지만 4개의 토픽과 모두 연결되는 가장 중요한 키워드가 된다.

그림 5. 대통령의 해양관 토픽 네트워크



네트워크를 보면 중심성이 있는 단어는 “해양”이며 모든 토픽을 연결시키는 중요한 단어이며 다음으로 “바다”가 3개의 토픽과 연결되어 있다. “세계”와 “산업”은 주변 토픽을 연결시키는 요소로 파악되고 있다.

결국 역대 대통령의 공통된 해양에 대한 인식은 바다를 중심으로 세계를 연결시키는 산업으로서 경제적으로 국가 미래에 중요하며 환경, 해운, 협력, 안보로서 해양이 중요하다라는 인식을 하고 있는 것으로 해석된다. 따라서 해양에 관한 대통령의 인식에 따라 행정부에서는 해양정책을 펼치는 방향을 제공하였다고 간주할 수 있다. 이것은 의원내각제

에 비하여 대통령 중심제가 행정부를 보다 강력하게 제어하기 때문에 가능하다.

하지만 정부의 운영은 실질적으로 행정부의 자료들에 의해 이루어지는 만큼 행정부처의 장관들의 인식과 대통령의 인식이 다를 수도 있다. 본 연구에서는 대통령의 연설문을 중심으로 해양관을 파악하였지만 정부의 해양 행정운영 방향을 동시에 파악하여 비교 분석한다면 보다 의미 있는 연구가 될 수 있다. 향후에 연구에서는 장관 연설문과 예산안 등의 성과지표를 비교분석한다면 계량적 연구의 성과로 나타날 수 있다.

## V. 결론

대통령의 연설은 정책의 방향을 가늠하고 또한 정부의 실천으로 옮겨지기에 어떠한 인식과 가치관을 가지고 있는지를 확인하는 것은 의미가 있다. 그간의 연구방법에서는 정성적이고 비정형화되어 있는 텍스트의 분석이 어려운 방법으로 알려져 있었으나 최근의 정보기술의 발전은 빅데이터 연구방법으로 진화되어 왔다.

특히 해양국가인 우리나라에서 대통령이 해양에 관해 어떻게 인식하고 있는지를 확인하는 것은 일반 국민뿐만 아니라 해운항만산업 관여하고 있는 국민들에게는 관심의 대상이 된다. 이에 국가적으로 해양산업의 중요성을 기념하는 “바다의 날” 기념식에 참석하는 대통령의 연설문을 1996년부터 2019년까지 24개년의 연설문을 분석하였다.

국가기록원에서 확보한 비정형 데이터인 연설문을 텍스트마이닝 가운데 토픽모델링이라는 분석방법으로 분석한 결과 대통령은 모두 자신의 국정운영 방향에 부합되는 해양관을 가지고 연설문이 발

표되었음을 확인하였다. 해양의 고유가치인 해양-산업-자원 토픽은 훼손되지 않고 지속적으로 대통령이 모두 강조하고 있음도 확인되었다. 따라서 빅데이터 분석방법으로 비정형화된 대통령의 연설문을 분석하더라도 유효한 분석방법으로 사용할 수 있음을 발견하였다.

분석결과 대통령의 당시 재임시에 강조하였던 국정운영의 주요 방향과 유사하였다. 행정부의 수반을 겸하는 대통령 중심제 정치체제에서 대통령의 인식이 담겨있는 연설문과 국정운영 방향과 같다는 것은 정부운영의 안정성을 제공한다는 점에서 긍정적인 인식이었다.

향후의 연구에서는 행정부와 비교하여 국정운영 방향과 대통령의 연설문이 일치하는 지를 비교분석한다면 의미가 있을 것이다.

### 참고문헌

각종 기념일 등에 관한 규정(대통령령 제27619호, 2016.11.30.시행).  
 국가기록원, 검색일자: 2019. 12. 10. <http://www.archives.go.kr>  
 국정기획자문위원회(2017), 문재인정부 국정운영 5개년 계획 및 100대 국정과제, 서울: 청와대.  
 김대중평화센터 <http://www.kdpeace.com> 검색일: 2020.3.20.  
 김동환(2000), 김대중 대통령의 인과지도: 1997년도 금융 위기의 원인과 극복에 관한 김대중 대통령의 시스템 사고, 한국 시스템 다이내믹스 연구, 제1권 제1호, 5-28.  
 김영삼민주센터 <http://kyscd.org> 검색일: 2020.3.20.  
 김혁(2014), 대통령의 리더십과 정책의제설정정양태에 대한 연구: 역대 대통령들의 주요 연설문에 대한 내용 분석을 중심으로, 한국정치연구, 제23집 제2호, 77-102.  
 대통령기록관, 검색일자: 2020. 2. 10. <http://pa.go.kr/research/index.jsp>

박준형·유호선·김태영·한희정·오효정(2017), 대통령 연설기록 내용분석을 통한 역대 대통령의 중심가치 연구, 한국기록관리학회지, 제17권 제2호, 57-78.  
 Cyram(2015), NetMiner 프로그램 백서: Introduction to NetMiner 4, 서울: 사이람.  
 이창길(2010), 정권초기의 가치지향과 정책우선순위: 참여정부와 이명박정부의 언어네트워크 비교분석, 한국행정학보, 제44권 제3호, 165-189.  
 김성국·이태휘(2019), 텍스트마이닝을 이용한 대통령의 물류관 분석에 관한 연구, 한국해운물류학회 2019년도춘계정기학술대회, 서울 중앙대학교, 1-26.  
 장정우·최경호(2012), 언어네트워크분석을 이용한 통계법 내용분석, 통계연구, 제17권 제2호, 53-66.  
 차두원·유지연(2013), 창조경제 개념과 주요국 정책분석, 서울: 한국과학기술기획평가원.  
 최원기(2013), 이명박 정부의 에너지 외교 정책: 평가와 정책적 시사점, 국가안보와 전략, 제13권 제3호, 77-108.  
 한겨레신문(2003.2.25.), 취임사로 본 '동북아 중심국가론'. 한신잡(2015), 빅데이터와 사회과학하기: 자료기반의 변화와 분석전략의 재구상, 한국사회학, 제49집 제2호, 161-192.  
 Aggarwal, Charu C.(2018), Neural Networks and Deep Learning: A Textbook, New York: Springer International Publishing.  
 Anderson, Chris(2008), The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, WIRED, 16.07, <https://www.wired.com/2008/06/pb-theory/>  
 Blei, David M.(2012), Probabilistic top models, Communications of the ACM, 55(4), 77-84.  
 Blei, David M., Andrew Y. Ng, and Michael I. Jordan(2003), Latent dirichlet allocation, Journal of machine Learning research, 3. 993-1022.  
 Box, George(1987), Empirical Model-Building and Response Surfaces, New York: Wiley.  
 Campbell, Karlyn Kohrs and Jamieson, Kathleen Hall(1990), Deeds Done in Words: Presidential Rhetoric and the Genres of Governance, Chicago: University of Chicago Press.  
 Cobb, R., J. K. Ross, and M. H. Ross(1976), Agenda building as a comparative political process, American Political Science Review, 70(1),

- 126-138.
- Danowski, J.(1993), Network Analysis of Message Content, Progress in Communication Sciences, 12, 198-221.
- Economist(2015), The data deluge: Five years on, <http://www.veritas.com/content/dam/Veritas/docs/reports/EIU-veritas-data-deluge.pdf>(검색일: 2020. 2. 29.).
- Halevy, A., Norvig, P. and Pereira, F.(2009), The Unreasonable Effectiveness of Data, IEEE Intelligent Systems, 24(2), 8-12.
- Hearst, Marti A.(1999), Untangling Text Data Mining, Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20-26, 3-10.
- Howlett, Michae and Ramesh, M.(2003), Studying Public Policy: Policy Cycles and Policy Subsystems, 2nd edition, Toronto: Oxford University Press.
- Hume, David(1978), A Treatise of Human Nature, 2nd edition, Oxford: Oxford University.
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M.(2009), Computational Social Science, Science, 323, 721-723.
- Linz, J.(1985), Democracy: Presidential or Parliamentary does it make a difference?, Connecticut: Yale university press.
- Marini, Margaret M. and Singer, Burton(1988). Causality in the Social Sciences, Sociological Methodology, 18, 347-409.
- May, Peter J.(1991), Reconsidering Policy Design: Policies and Publics, Journal of Public Policy, 11(2), 187-206.
- Scott, John(2000), Social Network Analysis: A Handbook, 2nd edition, London: SAGE Publications.
- Song, J., Kim, M., and Bhattacharjee, A.(2014), Social Science Research: Principles, Methods, and Practices [In Korean], KOCW Open Access Textbooks, Daegu, Korea: Korea Open Courseware.
- Steyvers, M., & Griffiths, T. (2007), Probabilistic topic models, in Landauer, T. K., McNamara, D. S., Dennis S. and Kintsch, W. (Eds.), Handbook of latent semantic analysis, Lawrence Erlbaum Associates Publishers. 427-448.
- NetMiner 4(Ver.4.4)<http://www.netminer.com>

# 텍스트마이닝을 이용한 한국 대통령의 해양관에 관한 연구

김성국 · 이태휘

## 국문요약

대통령 중심제의 정치체제에서는 대통령의 언어가 국가정책의 형성과 의사결정 과정에 지대한 영향을 미치게 된다. 대통령의 이념과 중심가치에 따라 정책우선순위가 결정되고 그 우선순위에 따라 다양한 정책이 수립되고 집행된다. 그래서 대통령의 연설문을 내용분석하는 연구가 관심의 대상이 되고 있다.

대통령의 연설문은 언어 자료이기 때문에 비정형이면서 비구조화 된 텍스트를 분석하기 위해서는 기계학습과 딥러닝의 방법을 통해 빅데이터 분석이 이루어지고 있다. 본 연구에서는 1996년부터 24년간에 걸쳐 “바다의 날” 기념식의 대통령 연설문을 확보하여 텍스트마이닝 방법의 일종인 토픽모델링의 방법으로 분석하였다.

분석결과 역대 대통령은 모두 자신의 국정운영 방향에 부합되는 해양관을 가지고 연설문을 발표하였다는 것을 확인하였다. 해양의 고유가치인 해양-산업-자원 토픽은 훼손되지 않고 지속적으로 역대 대통령이 모두 강조하고 있음을 확인하였다.

주제어: 텍스트마이닝, 연설문, 토픽모델, 내용분석

