

텍스트마이닝을 활용한 북한 지도자의 신년사 및 연설문 트렌드 연구

박 철 수*

Discovering Meaningful Trends in the Inaugural Addresses of North Korean Leader Via Text Mining

Chul-Soo Park*

Abstract

The goal of this paper is to investigate changes in North Korea's domestic and foreign policies through automated text analysis over North Korean new year addresses, one of most important and authoritative document publicly announced by North Korean government. Based on that data, we then analyze the status of text mining research, using a text mining technique to find the topics, methods, and trends of text mining research. We also investigate the characteristics and method of analysis of the text mining techniques, confirmed by analysis of the data.

We propose a procedure to find meaningful tendencies based on a combination of text mining, cluster analysis, and co-occurrence networks. To demonstrate applicability and effectiveness of the proposed procedure, we analyzed the inaugural addresses of Kim Jung Un of the North Korea from 2017 to 2019. The main results of this study show that trends in the North Korean national policy agenda can be discovered based on clustering and visualization algorithms. We found that uncovered semantic structures of North Korean new year addresses closely follow major changes in North Korean government's positions toward their own people as well as outside audience such as USA and South Korea.

Keywords : Text Mining, North Korea, Addresses, Leader, Text Analysis

1. 서 론

4차 산업혁명 시대에 접어들면서 데이터가 미래 발전의 핵심 성장 동력으로 급부상 하고 있으며 그 중요성과 활성화 방법에 대한 연구와 논의가 활발하게 진행되고 있다. 최근에는 인공지능(AI), 클라우드컴퓨팅, 사물인터넷(IoT), 빅 데이터 분석 등 디지털 비즈니스의 확대로 데이터가 홍수처럼 생성됨에 따라 이를 처리하기 위한 분석기술과 전문가들의 중요성이 더욱 증가하고 있다.

글로벌 시장조사 기관인 IDC에 의하면 2025년이 되면 전 세계의 데이터 생산량은 약 163ZB 정도의 크기가 될 것으로 예측하였고, 2016년에 생성된 16.1ZB에 비하면 약 10배정도 늘어난 규모이다. 디지털기술의 발전과 함께 등장한 스마트 및 모바일 기기의 확산과 이로 인한 소셜네트워크 서비스(SNS)의 활성화, 그리고 IoT의 보급 등으로 인해 빅 데이터의 활용과 중요성은 더욱 증가하고 있다.

세계적인 IT 리서치 전문회사인 가트너의 수석부사장 피터 손더가드(Peter Sondergaard)는 “빅 데이터는 21세기의 원유”라고 정의한 바 있다.¹⁾ 원유를 어떻게 가공하느냐에 따라 만들어낼 수 있는 제품의 종류가 무궁무진하듯이, 빅데이터 역시 데이터를 어떻게 활용하고 분석하느냐에 따라 이전에 사고하지 못하고 발견하지 못했던 수많은 분야에서 새로운 가치들을 창조해낼 수 있을 것이다. 이처럼 방대한 빅 데이터의 발생이 개인의 일상뿐 아니라 유통, 제조, 서비스, 방송, 언론, 패션, 바이오, 보건·의료, ICT 등 다양한 분야에서 축적되면서 데이터에 대한 분석이 점점 더 중요해지고 있다.

데이터에서 유용한 정보를 발견하는 과정을 데이터 마이닝(Data Mining)이라고 하며, 이러한 데이터 마이닝 과정을 통해 가치 있는 정보를 만들어 의사결정에 적용할 수 있게 된다. 최근 빅 데이터 분석 기술이 발달하여 기존에 많은 연구자들이 숫자형태의 데이터로 대표되는 정형화된 데이터(Structured Data)의 분석을 넘어 이제는 텍스트, 동영상, 음성, 이미지 등과 같은 비정형화된 데이터(Unstructured Data)에 대한 연구도 활발하게 진행하고 있다.

이러한 연구가 정치·경제·문화·사회적으로 의미를 가지는 다양한 분야에 적용이 되고 있다. 더욱이 오늘날 인터넷 및 스마트 폰 등을 통해 생산되는 SNS 데이터의 양이 폭발적으로 증가하면서 이에 대한 분석의 필요성이 계속 증가되고 있다. 이런 데이터는 실시간으로 생성되고 비교적 짧은 길이를 가지며, 구조화되지 않고 비문법적인 문장 구조의 특성을 가지며, 이를 위해 소셜 네트워크 분석과 같은 다양한 분석 기법을 이용하여 언론과 커뮤니케이션, 그리고 정치 분야에서 많이 활용되고 있다. 또한 텍스트 문서 등에서 객관적인 글을 요약하거나 분석하는 일반적인 텍스트마이닝과 달리, 오피니언 마이닝, 즉 감성 분석을 통해 개인의 감성을 수집하여 이를 정량화하는 분석을 시도함으로써 사회과학 및 정치 커뮤니케이션 분야에서도 다양한 데이터를 통한 체계적, 객관적, 과학적 분석의 가능성이 열리게 되었다(Hymans, 2008; Rich, 2012). Cho et al.(2015)은 텍스트 트렌드 분석 방법을 통해 미국 대통령 취임사의 시계열 통합 트렌드를 도출하는 연구를 수행하였다.

Park et al.(2015) 연구에서는 1946년부터 2015년까지 발표된 북한 신년사를 전수분석 하여 북한 신년사 텍스트가 가진 특성을 고려하여 신년사의 거시적 구조와 문서 간 상관성을 중심으로 분석하는 거시적 분석과 특정 단어의 위치와 빈도, 관련어와의 결합방식을 분석하는 미시적 분석으로 나누어 연구를 수행하였다. 대통령 등 국가지도자의 연설문이나 메시지는 국가적 여러 당면 과제와 사회적인 방향성이 기록된 자료로서, 역사적인 가치가 매우 크며 시대변화에 따르는 국정운영의 추이를 살펴볼 수 있다는 점에서 이러한 연설문이나 국가적인 메시지들은 효용가치가 매우 높다고 할 수 있다(Cho et al., 2015). 현재까지 수행된 대부분의 연구는 크게 연설문에서 사용되는 특정 단어의 추이를 관찰함으로써 트렌드를 도출하거나(Lim, 2002), 연설문 내용에 포함된 국가적 주제 즉, 정치, 경제, 사회, 복지, 통일, 외교 등의 측정지표를 연구자가 설정하고, 정량화하여 그 추이를 분석하였다(Kim, 2014). 그 외에, 연설문에 포함되는 단어들의 관계를 활용하여 네트워크를 형성하고 분석하는 연구가 있었다(Kim, 2013).

본 연구에서 분석대상으로 설정한 북한지도자의 신년사와 연설문은 북한의 당면 과제와 정치, 경제,

1) <http://www.dailybizon.com/news/articleView.html?idxno=13108>(검색일 : 2019년 4월 20일).

사회적인 변화와 지향점이 기록된 자료로서, 그 역사적인 가치가 크며, 시대 상황적인 정세가 모두 집대성되어 있다. 따라서 북한지도자의 신년사와 연설문을 활용한 트렌드 연구는 시대변화에 따른 국정운영의 추이를 살펴볼 수 있다는 점에서 그 효용가치가 높다고 판단할 수 있다. 북한에서 지도자의 신년사는 한 해 동안 북한이 나아가 할 분야별 과업과 수행방도를 밝히는 매우 중요한 문건으로 외무성 등 모든 중앙 부처들은 신년사에 반영된 내용을 구체적으로 어떻게 관철하겠다는 분야별 '작전전술안'을 문건으로 만들어 김정은에게 제출하며, 매해 6월이 되면 상반기 신년사 관철총화, 12월이 되면 부서별로 직능총화라는 것을 하여 신년사에 반영된 내용을 어떻게 관철했는가를 평가 한다.²⁾

특히 2018년 북한지도자의 신년사를 기준으로 그 이전과 이후의 남·북·미 관계가 급격히 변화하였다. 그래서 그 이전과 이후, 그리고 최근의 변화를 포함하는 북한지도자의 신년사와 연설문을 분석한다는 것은 한 국가집단의 지향점과 정세를 파악하고 이를 바탕으로 앞으로의 남·북·미 관계를 살펴보는 데 커다란 의의가 있을 것으로 판단된다.

그래서 본 연구에서는 다음과 같은 몇 가지의 목적을 가지고 연구를 수행하고자한다. 첫 번째 목적은 최근의 남·북·미 관계의 극적인 변화를 반영하여 기간을 나누어 각 기간의 북한 지도자의 신년사와 연설문을 텍스트마이닝 기법을 적용하는 것이다. 그 기간은 2018년 이전을 '평화 조성 이전기', 2018년 1월부터 2019년 2월 하노이 회담 전 까지를 '평화 조성기', 그리고 2019년 2월 이후를 '평화 조성 침묵기'로 구분하였다. '평화 조성 이전기'와 '평화 조성기'는 매년 초에 발표하는 김정은 위원장의 신년사가 분석대상이며, 2019년 2월 하노이 회담 이후 '평화 조성 침묵기'에는 4월에 열린 북한 최고인민회의의 김정은 위원장의 연설문을 분석 대상으로 하였다. 북한 헌법에 따르면 최고인민회의는 "최고주권기관"으로 남한의 국회에 해당하는 입법부로서 제14기 1차 회의가 2019년 4월 11일 평양에서 열렸다. 특히 하노이 2차 북-미 정상회담 합의 무산 이후 한반도 정세의 긴장이 높아지는 상

황에서 정세 흐름과 관련한 최대 관심사는 '하노이 회담' 이후 북한의 정리된 김정은 국무위원장의 대외 메시지이기 때문이다.

두 번째 목적은 앞서 제시한 기간에 있었던 북한 지도자의 신년사와 연설문을 텍스트마이닝 분석기법을 통하여 신년사와 연설문에 담겨진 단어들의 트렌드와 토픽의 변화과정을 살펴보는 것이다. 김정은 위원장의 신년사와 연설문이 최근의 한반도 및 남·북·미관계의 변화를 신년사와 연설문의 내용에 담겨있는 단어분석을 통해 트렌드가 어떻게 변화하고 있는지 연구를 수행하였다.

이를 위해 본연구의 구성은 다음과 같다. 제 2장에서는 선행 연구들에 대한 내용을 정리하였으며 3장에서는 본 연구에서 활용된 북한지도자의 연설문에 대한 연구방법을 위해 연구수행절차, 자료의 수집, 키워드 추출 및 키워드 전처리 방법을 설명 할 것이다. 4장에서는 북한지도자의 각 년도 별 신년사 분석결과를 보여주고 이를 종합하여 트렌드를 보여줌으로서 북한 지도자의 연설문 속에 그 시대의 상황과 변화 그리고 북한의 상태와 남·북·미 관계를 가늠해 볼 수 있을 것이다. 제 5장에서는 본 연구의 결론 및 기대효과와 함께 본 연구의 한계점에 대하여 논의하고 향후 연구 방향을 모색하였다.

2. 선행연구

텍스트마이닝은 자연어 처리 기술을 기반으로 직접적인 연관을 보여주지 않는 비정형 텍스트에서 숨겨진 관계 또는 패턴을 도출하여 의미 있고 활용 가치가 높은 정보 또는 지식을 창출하는 기법이다[Judita, Stevenson, and Gaizauskas, 2015; Ronen and Daga, 1995]. 텍스트마이닝을 위한 분석방법에는 토픽트래킹(Topic Tracking), 소셜네트워크 분석(Social Network Analysis), 감성 분석(Sentiment Analysis)[Liu, 2012], 오피니언 마이닝(Opinion Mining), 정보 추출(Information Extraction), 텍스트 분류(Classification), 텍스트 군집화(Clustering) 등이 있다[Fan, et al., 2006]. 텍스트 분석에서 문서 중 단어의 중요도를 측정하는 방법은 특정한 단어가 한 문서 내에서 얼마나 자주 반복되는지, 그리고 문서 그룹 내에서 동일한 단어가 얼마나 많이

2) [https://thaeyongho.com/2019/01/02/김정은-신년사-평가-2-북한에서-신년사가-가진-의미/\(2019년 4월 20일 검색\).](https://thaeyongho.com/2019/01/02/김정은-신년사-평가-2-북한에서-신년사가-가진-의미/(2019년 4월 20일 검색).)

출현하는지를 측정 할 수 있다. 텍스트마이닝의 중요한 장점은 텍스트 내에서 실제적이고 잠재적으로 이루어지고 있는 메인의 논의 및 표현 내용이 무엇인지를 발견하는 것이다[Chakraborty et al., 2013]. 즉 텍스트를 어떤 기준에 따라 정해진 범주로 분류하는 것 [Chen and Chen, 2011], 유사한 텍스트들의 군집 군을 발견, 텍스트 안에서 특정 단어들 간의 특별한 규칙 발견, 문서 요약[Kim et al., 2013], 정보검색 [Pai et al., 2013] 등을 말한다. 특히 텍스트 안에서 특정 단어들 간의 특별한 규칙을 발견하기 위해 다양한 길이의 텍스트 데이터를 분석 하면서 특정 단어 (형태소)가 반복적이며, 인접하여 나타나는 동시출현 (Co-Occurrence) 빈도와 형태를 이용해 단어의 영향력을 다양한 네트워크 형태로 보여주며 이를 근거로 단어들 간의 연관성을 보이는 단어들 간의 관계를 보여줌으로써 의미론적 연관구조를 도출해 낼 수 있다. 또한 소셜네트워크 분석에서는 사회구조와 상호의존성 그리고 개인, 그룹, 조직의 작업패턴을 분석할 때 유용하며, 다양한 원천 데이터를 수집하고 관계를 분석하며, 그 관계의 품질, 효과, 영향도 등을 포괄적으로 분석하여 나타내는 것이다.

이 외에도 Myung et al.[2008]은 상품리뷰를 분석하는 시스템을 구축하고 시스템을 활용해 상품순위를 예측하였고, Hahm and Lee[2016]은 비정형데이터의 수집을 통해 가치를 창출하는 빅데이터 비즈니스 모델에 대해 분석하였다. 또한 Kim et al.[2012]은 뉴스 텍스트 데이터의 분석을 통해 투자의사결정 모형을 개발하였으며, Ryu et al.[2016]은 텍스트마이닝을 기반으로 하는 주식 투자전략 수립하여 우수한 성과를 입증하였다. 그리고 많은 연구들이 사람들의 의견이 표출된 텍스트로부터 특정 감정을 추출하고자 하였다 [Liu, 2010; Narayanan et al., 2009; Sadamitsu et al., 2008] 학술지의 논문 주제어간 연관관계 연구 [Cho and Kim, 2012; Kim et al., 2016], 지속가능보고서에 나타난 CEO 메시지의 언어적 구문적 특성을 다양한 관점에서 주관적 객관적인 지속가능성 성과와 관련한 실증연구[Yook, 2018], 국내 산업공학 연구 기법의 트렌드 분석[Cho et al., 2014], 특허문서를 활용한 기술 추이 연구[Kam et al., 2013; Kim et al., 2009; Park et al., 2014], 인터넷 검색추세를 활용한 주식투자전략 연구[Kim and Koo,

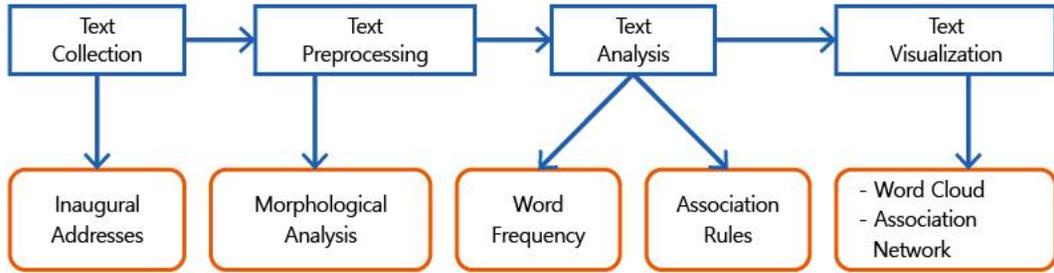
2013] 등 다양한 연구들이 있다. Tumasjan et al. [2010]은 트위터 메시지에 실제 정치의 여론이 잘 반영되어 있는지 보기 위해 2009년 독일 연방 선거를 대상으로 텍스트마이닝 기법을 사용하였다.

또한 텍스트데이터의 트렌드 연구는 기존의 전문가 의견, 설문조사 등과 같은 전통적인 방법에서 벗어나, 다양한 분야에서 발생하는 텍스트데이터 자체를 분석 연구함으로써 데이터 기반의 객관적인 결과를 탐구하는데 연구의 의의가 있을 수 있다. 특히 연설문 내용에 포함된 국가적 주제 또는 뉴스와 언론보도, SNS 등 정치, 경제, 사회, 복지, 통일, 외교 분야에서 발생하는 텍스트데이터를 이용하여 이를 정량화하고 그 추이를 분석하는 연구들도 많이 있다[Kim, 2014; Lee and Kim, 2018]. 그 외에, 연설문에 포함되는 단어들의 관계를 활용하여 네트워크를 형성하고 분석하려는 연구[Kim, 2013], 텍스트마이닝을 이용한 김정은 정권의 대남정책 및 통일담론을 분석하는 연구[Oh and Lee, 2016], 텍스트마이닝을 이용한 북한 보도 동향과 북한 도발과의 연관성 분석[Lee and Moon, 2016], 북한 신년사(1946~2015)에 대한 자동화된 텍스트 분석 연구[Park, et al., 2015], 텍스트마이닝을 활용한 미국 대통령 취임 연설문의 트렌드 연구 [Cho et al., 2015] 등 국가 지도자들의 연설문의 트렌드 분석을 통한 국가의 지향점을 살펴보는 연구들이 진행되고 있다.

3. 연구 방법

3.1 연구 수행 절차

본 연구는 북한 지도자의 신년사 및 연설문을 활용하여 텍스트마이닝 방법론을 적용하여 분석하였다. 본 연구의 대상이 되는 북한 지도자의 신년사 및 연설문에 담겨진 단어와 토픽들을 통하여 최근 한반도의 안보 및 평화의 상태를 가늠해보기 위한 연구로서 평화 조성 이전기(2017년 신년사), 평화 조성기의 신년사(2018년, 2019년 신년사) 그리고 2019년 2월 하노이 북미 정상회담 이후 평화 조성 침묵기(2019년 최고인민회의 제14기 시정 연설) 이상 3개의 시점으로 나누어 분석하였다. 북한 지도자의 연설문은 국정운영의 기본방향과 청사진을 기록한 자료로서 텍스트마이닝 방법을 이용하여 북한지도자의 연설에 담긴 내용



〈Figure 1〉 Research Phases and Methods

변화와 언어 활용의 차이점을 유추함으로써 북한의 정책의 기본 방향을 가늠해보는 연구가 될 것이다. 키워드 분석 및 토픽분석이 본 연구의 주제 및 분야에 활용 될 것이며, 〈Figure 1〉는 연구단계 및 방법을 정리한 것이다. 〈Figure 1〉은 Kim et al.[2012] 연구를 참고하여 저자가 재작성 한 것이다.

본 연구를 위해 분석대상인 4개의 신년사 및 연설문을 살펴보면 2017년 신년사의 총 사용 단어 수는 10,422개 이고, 2018년 신년사의 총사용 단어 수는 11,712개, 2019년 신년사의 총사용 단어 수는 12,774개, 2019년 4월 시정연설의 총사용 단어 수는 11,958개이다. 본 연구의 대상되는 북한 지도자의 신년사 및 연설문은 북한의 일반 대중에게 까지 미디어나 신문, 텔레비전, 또는 문서형태로 내용이 전달되어진다. 그 내용과 문장을 분석 해보면 시간이 지날수록 보다 평이하고, 간결한 형태를 보이고 있음을 유추해 볼 수 있다. 그러나 본 연구에서는 이런 단편적인 추이 발견을 넘어 텍스트 마이닝 방법론을 적용하여 보다 포괄적이고 세밀한 신년사 및 연설문의 트렌드 분석을 수행하였다. 신년사 및 연설문의 문서의 특성을 대표하는 객관적인 주요 키워드를 추출하고, 키워드 빈도 수 와 이를 바탕으로 문서-단어 교차표를 생성하기 위한 전처리 작업을 수행하였다. 그리고 모든 문서 내에 포함된 단어를 추출하고, 불용단어를 제거한다. 불용단어는 의미 없는 단어들의 집합이며 관사, 전치사, 조사, 접속사, 의미 없는 숫자 등이 그 대상이 된다.

본 연구를 위해서 사용한 데이터 분석 프로그램은 'R' 프로그램으로 데이터분석을 위한 오픈소스이다. 오픈소스는 누구나 어디서나 사용가능한 프로그램을 말한다. R은 통계분석, 데이터 시각화, 데이터 마이닝, 빅 데이터, 인공지능 등에 강점을 가지고 있고, 다양한 전공 분야에서 활발히 이용되고 있다. 이 프

로그램에는 패키지(Package)라는 추가 기능이 있고, 현재 약 1만 4천여 개가 넘는 패키지로 구성되어 있다.

3.2 데이터 수집과 전 처리

본 연구에서 불용어 처리를 위해 R프로그램에서 제공하는 Natural Language Toolkit에서 corpus의 불용어 사전을 활용하여 불용어를 제거하였다. 그 후 문서 내에 포함된 기호 및 무의미한 단어는 연구자가 확인하고 삭제했는데, 이때 제거된 글과 기호는 '들', '당', '비롯', '하기', '시', '-', '?', '2' 등과 같은 것이다.

본 연구의 분석대상인 4개의 신년사 및 연설문의 사용 단어 수는 년도 별로 2017년 신년사의 총 사용 단어 수는 10,422개 이고, 2018년 신년사의 총사용 단어 수는 11,712개, 2019년 신년사의 총사용 단어 수는 12,774개, 2019년 4월 시정연설의 총사용 단어 수는 11,958개이고 핵심어를 추출하기 위하여 과정으로 TF-IDF 분석을 수행하였다. TF-IDF는 여러 문서로 이루어진 문서의 집합이 존재할 때, 각각의 문서에 포함된 단어의 중요도를 산출 하는 통계적인 수치로써, 문서 내 단어의 출현 수를 나타내는 TF와 총 문서에서의 단어의 출현비율의 역수를 취한 IDF를 활용하며, 식 (1)과 같이 계산된다.

$$TF-IDF_i = TF_i \times IDF_i = TF_i \times \log_2 \frac{N}{n_i}, (1)$$

여기서 TF_i 는 단어 i 가 모든 문서에서 관찰되는 양이며, DF (문서 빈도, Document Frequency)라고 하며, 이 값의 역수를 IDF (역문서 빈도, Inverse Document Frequency)라고 한다. 따라서 단어의 중요도를 나타내는 TF-IDF 값은 TF와 IDF를 곱한

값으로 점수가 높은 단어일수록 다른 문서에는 많지 않고 해당 문서에서 자주 등장하는 단어를 의미하며, 특정 단어의 출현양이 많을수록 증가하지만, 모든 문서에서 빈번하게 사용되는 단어는 그 정도에 따라 값이 낮게 계산된다.

3.3 분석 방법

본 연구의 분석대상인 4개의 신년사 및 연설문의 빈도분석을 실시하였다. 그리고 형태소 분석에는 R의 KoNLP패키지에서 제공하는 SimplePos09 형태소 분석기를 사용하였다. 형태소분석을 통해 단어를 추출하는 과정을 텍스트 분석에서는 전처리(Preprocessing)라고 부른다. 전 처리된 자료는 해당 연도를 문서 값으로 갖는 문서(Document)로 저장되는데, 이 문서가 텍스트 분석을 위한 분석단위이다.

연관성 분석은 항목 간의 상호 관계를 분석하는 것이다. 텍스트 분석에서는 단어와 단어의 상호 관계를 분석하기 위하여 단어의 동시발생(Co-Occurrence)을 분석한다. 동시발생이란 한 문장, 문단 또는 텍스트 단위에서 같이 출현한 단어가 자주 발생할 확률을 나타내는 언어학 용어이다. 연관성 분석의 측도는 지지도(Support), 신뢰도(Confidence)와 향상도(Lift) 값을 잘 보고 결정해 한다. 지지도란 전체 문서 중 단어 A와 단어 B가 동시에 발생하는 정도를 나타낸다. 신뢰도는 단어 A를 포함한 문서 중에서 단어 A와 단어 B가 함께 발생할 확률이 어느 정도인가를 나타낸다. 향상도는 단어 A가 발생하지 않았을 때 단어 B가 발생할 확률에 비해 단어 A가 발생하였을 때 단어 B의 발생 확률 증가 비율이다. 이 개념에서 출발한 동시출현 네트워크(Co-Occurrence networks)는 특정 텍스트 단위에서 공동으로 출현한 단어의 집합적 상호 연결을 표현하는 방식이다. 시각화를 위해 R의 qgraph와 networkD3 패키지를 활용하였다.

워드 클라우드(Word Cloud)는 대표적인 텍스트 시각화기법 중에 하나로 데이터의 연결과 그룹화를 표현하는 데 초점을 둔다. 워드 클라우드는 최소의 의미를 지니는 문장 구성 성분인 형태소를 분석하고 그 빈도에 따라 문자의 크기를 결정한다. 이러한 시각화는 텍스트에서 키워드의 빈도를 직관적이고 빠르게

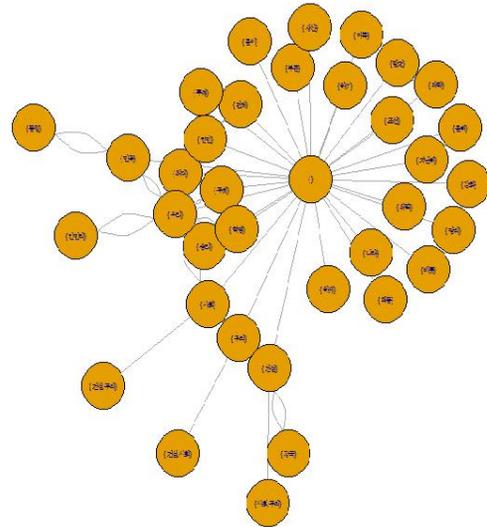
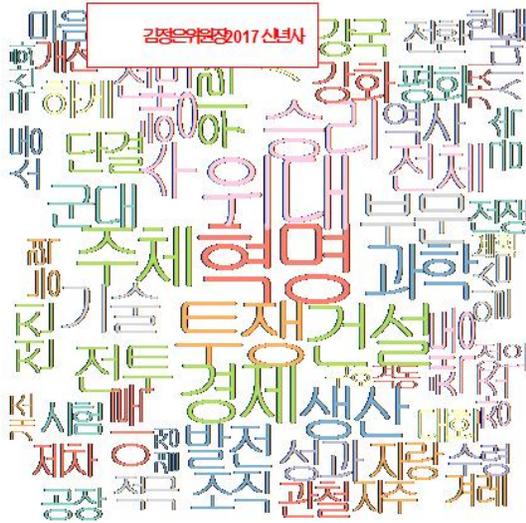
인지할 수 있는 장점이 있다. 이를 위해 word cloud와 igrph 패키지를 사용하였다.

4. 북한 지도자 신년사 및 연설문 분석결과

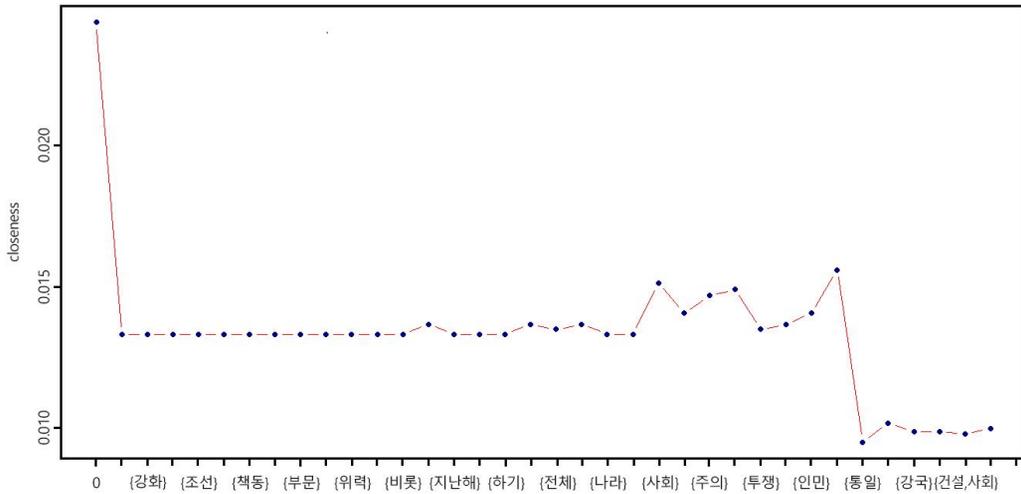
4.1 2017년 북한지도자 신년사 분석

2017년 북한 지도자의 신년사에 몇 가지의 특성을 지니고 있다. 먼저 선대(先代)와는 차별화된 김정은 시대의 본격적인 시작을 의미하는 것이다. 그래서 선대에 대한 언급의 빈도가 2013년 25회, 2014년 11회에 걸쳐 나타났던 김일성, 김정일 명칭은 2015년에는 사라졌으며, '장군' 혹은 '수령' 등의 우회적 표현만이 등장하였고, 2016년에는 전현 언급이 없었으며, 2017년에는 선대를 언급하는 표현이 완전히 사라졌다. 둘째는 김정은 식의 새로운 단어 명명이 두드러지게 나타나기 시작 하였다. 김정일의 국가 비전이라 할 수 있었던 '강성대국' 혹은 '강성국가' 대신 '사회주의강국'이란 단어가 5회 등장하고 '자강력' 혹은 '자강'이란 표현이 새롭게 나타나기 시작하였다. 2016년에 처음 등장한 '자강'의 단어가 2017년 신년사에서 5회로 증가하였다. 셋째는 신년사에서 '병진정책'의 표현이 사라지기는 하였지만, '핵무력'의 강화를 위한 의지는 도처에서 나타났다. "우리식의 위력한 주체 무기들을 더 많이 개발·생산하여 선군혁명의 병기창을 억척같이 다져야 합니다"라고 언급한 부분이나, "...핵무력을 중추로 하는 자위적 국방력과 선제공격 능력을 계속 강화해나갈 것입니다"란 역설은 이의 대표적인 사례라 할 수 있다. 넷째는 <Figure 2>에 Word Cloud에서 보는 것처럼 인민, 혁명, 경제, 주체, 투쟁, 과학, 승리, 민족 등이 주요 빈도수를 차지하고 있다.

2017년은 전반적으로 남북 관계가 매우 암울한 상황에서 출발할 수밖에 없는 여건이었으며, 이는 김정은의 신년사에서도 그대로 나타났다. 김정은은 한국에 대화나 협상 보다는 앞으로 호의적인 정부가 탄생하기를 기다리면서 당시의 정부 보다는 차이를 목표로 하는 메시지를 계속 보내고 있었다. 일부 평화공세가 있기는 하였지만 오히려 정치적인 제스처의 성격이 강하며, 도발적 행태 역시 한·미의 정책공조 틈새가 극대화되기를 기다리고 있었던 시기였다.



<Figure 2> Word Cloud of North Korean New Year Addresses, 2017 <Figure 3> Co-Occurrence Network of North Korean New Year Addresses, 2017



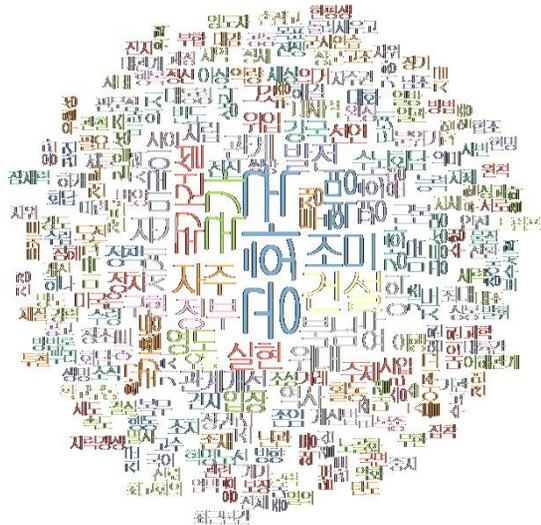
<Figure 4> Closeness of North Korean New Year Addresses, 2017

4.2 2018년 북한 지도자 신년사 분석

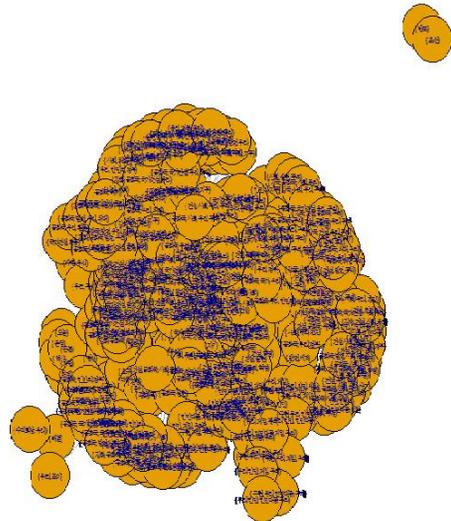
2018년 신년사는 평화수호와 억지력 보유에 초점을 맞춘 ‘핵 무력’ 언급과 완성을 명분삼아 남북관계에 많은 비중을 할애하며 적극적인 화해 제스처로 국면전환과 ‘평화를 사랑하는 책임 있는 핵강국’ 이미지를 전달하는데 주력 하였다. ‘핵 무력 완성’의 역사적 대업 성취 이후 갖게 된 ‘불가역적’인 전쟁억제력과 ‘전략국가’의 지위를 토대로 경제성과를 올리고 남북관계 개선

을 통해 남북 간 군사적 긴장해소 및 평화적 환경을 조성하겠다는 메시지로 구성되었다. 북한 정권수립 70주년을 맞아 평창동계 올림픽을 ‘민족적 대사’로 표현하고 민족의 위상을 과시하는 좋은 계기라는 의미부여를 하면서 성과를 기대한다고 하였다.

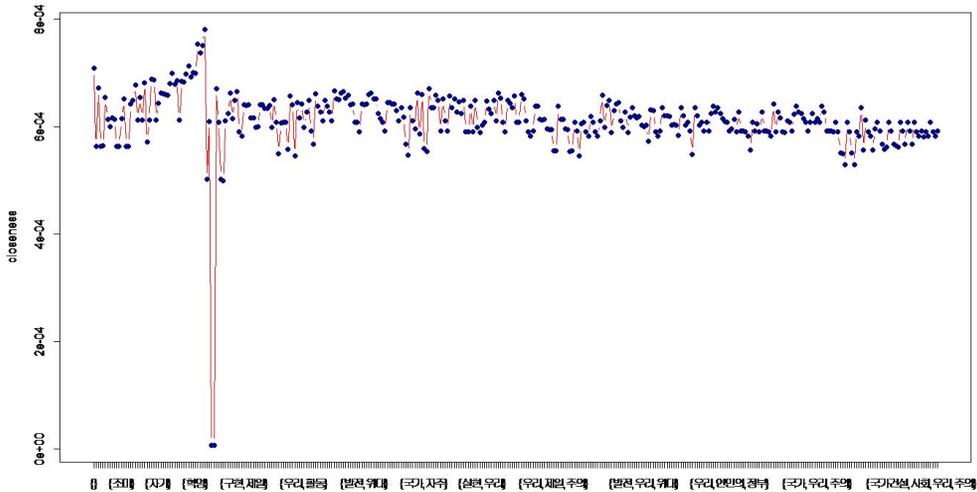
2018년 북한 지도자는 ‘핵 우선주의’를 표방하면서 ‘경제’와 ‘우리민족’에 초점을 맞추고 있다. 키워드 중심으로 보면 ‘핵’, ‘미국’, ‘전쟁’, ‘평화’ 등이 지난해에 비해 월등히 많은 것으로 조사되었다. 구체적으로 ‘핵’은



<Figure 11> Word Cloud of North Korea's Political Bureau Leadership's Speech, 2019



<Figure 12> Co-Occurrence Network of North Korea's Political Bureau leadership's Speech, 2019



<Figure 13> Closeness of North Korea's Political Bureau leadership's Speech, 2019

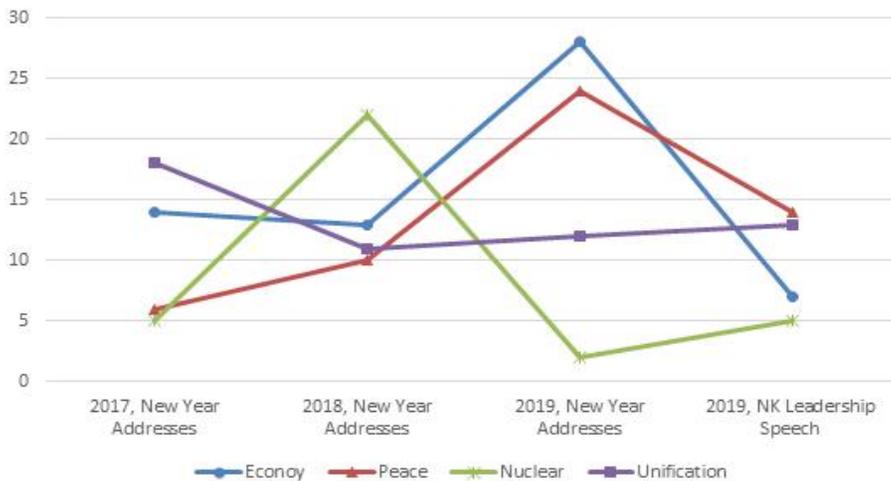
불수 있다. <Figure 11> Word Cloud에서 보는 것처럼 '공화국', '국가', '정부', '조미', '북남', '건설' 등이 강조되어 북한의 사회주의 국가 건설과 조미, 북남관계 개선을 의미하는 것 같다. <Figure 12>와 <Figure 13>에서 연설문 Co-Occurrence Network을 보면 다른 년도의 네트워크 모양과 다르게 네트워크가 분산이 안 되고 집중이 된 것은 같은 단어들이 집중되어 사용되고 그 단어들 간의 근접성이 매우 높다는 것을 의미한다.

4.5 종합 트렌드 분석

<Table 1>은 2017년부터 2019년 4월까지 북한 지도자의 신년사와 연설문의 키워드 빈도수를 보여 주는 것이다. '우리', '인민', '사회주의' 세 단어는 점차 사용빈도수가 해가 거듭될수록 증가하고 있다. 이는 김정은 위원장이 정상국가의 지도자라는 이미지를 구축하기 위한 것으로 판단된다. 또한 2019년 하노이 회담 결렬이후 4월 연설에 나타난 '우리', '인민', '사회주의',

<Table 1> Frequency of Key Word on North Korean New Year Addresses and Speech

	2017, New Year Addresses		2018, New Year Addresses		2019, New Year Addresses		2019, Political Bureau leadership's Speech	
	Key Word	Frequency	Key Word	Frequency	Key Word	Frequency	Key Word	Frequency
1	우리	37	우리	48	우리	75	우리	86
2	인민	29	인민	47	인민	54	인민	64
3	혁명	20	생산	25	사회주의	32	사회주의	40
4	생산	17	핵	22	평화	28	국가	38
5	통일	14	혁명	22	경제	28	공화국	37
6	경제	14	사회주의	20	자립, 자력	22	미국	26
7	힘	13	힘	20	투쟁	21	건설	24
8	사회주의	12	건설	20	힘	19	조미	21
9	투쟁	10	투쟁	19	국가	17	혁명	20
10	민족	9	민족	18	조선반도	16	북남관계	18
11	승리	9	승리	15	혁명	16	정부	17
12	위력	7	국가	14	북남관계	16	나라	17
13	주체	7	경제	13	생산	15	김일성	16
14	건설	6	전쟁	12	발전	14	힘	16
15	남조선	6	강화	11	민족	11	평화	14
16	사상	6	미국	11	조국	9	자주	13
17	일심단결	6	강국	10	통일	8	통일	12
18	조국	6	평화	10	강화	6	발전	11
19	평화	6	북남관계	9	공화국	6	세계	10
20	군대/핵	5	공화국	9	세계	6	투쟁	10



<Figure 14> Timeline Trends for Main Keywords of North Korean New Year Addresses and Speech

‘국가’, ‘공화국’ 등은 김정은 위원장의 체제 강화와 조선인민의 최고대표자로서 공고화를 의미하는 것으로 보인다. 또한 ‘김일성·김정일주의’라는 단어가 다수 등장하는 것은 김정은 시대의 새로운 통치이념 제시보다는 김일성·김정일주의 지속의지를 보여주는 것으로 김일성·김정일주의의 본질을 인민대중제일주의로 정식화 하는 것이다. 이는 노동신문, 1월8일자, 1월

21일자, 1월22일자에도 우리국가제일주의는 곧 김일성·김정일 조선 제일주의라고, 우리 국가 제일주의의 본질 중 하나로 인민대중제일주의라고 제시하였다. 그리고 “국가건설과 활동에서 자주 혁명 로선으로 철저히 관철시키고 혁명의 주체적 역량을 강화하며, 모든 분야를 우리 식으로 발전”시킨다는 연설내용에서도 키워드의 내용을 유추 해 볼 수 있다.

〈Table 2〉 Frequency for Main Keywords of North Korean New Year Addresses and Speech

Key words	Synonym	Before Peace Building	Peace Building Times		Silence Moment of Peace Building
		2017, New Year Addresses	2018, New Year Addresses	2019, New Year Addresses	2019, Political Bureau Leadership's Speech
Our	우리식, 우리당 우리민족, 우리힘 우리나라	37	48	75	86
Economy	건설경제력, 경제적 경제제재, 자립경제 국가경제, 경제발전 경제사업, 사회경제 경제구조, 지방경제 경제활성화	14	13	25	7
Peace	평화통일, 평화번영 평화적, 평화정착 평화체제, 평화보장 평화수호, 평화시대 평화지대	6	10	23	14
Nuclear	중핵, 핵무장력 핵시험, 핵위협 비핵화, 핵무기 핵전쟁, 핵무력 열핵무기, 핵강국 핵반격, 핵억제력 핵타격, 핵단두	5	21	2	5
Unification	조국통일, 통일적 평화통일, 통일방안 남북통일, 자주통일 통일운동, 통일대회 통일대진군	18	11	12	13

본 연구에서는 최근의 북한의 대내외적 정책변화를 특징짓는 주요 키워드를 ‘우리’, ‘경제’, ‘평화’, ‘핵’, ‘통일’로 단어들을 선정하였다. 그리고 이런 단어들이 신년사 및 연설문에서 주요 키워드로서 어떻게 변화하며, 그 추이가 어떻게 되는지를 〈Table 2〉와 〈Figure 14〉에서 나타내었다. 보는 것처럼 북한지도자의 신년사와 연설이 북한의 대내외적 정책변화에 따라 신년사와 연설문의 빈도수에도 영향을 미친다는 것을 알 수 있다.

먼저 2017년(평화조성 이전기) 신년사에서는 첫째 김정은식의 새로운 단어 명명이 두드러지기 시작했다. 김정일의 국가 비전이라 할 수 있었던 ‘강성대국’ 혹은 ‘강성국가’ 대신 ‘사회주의강국’이란 단어가 도처에 등장(5회)함으로써 김정은식 브랜드네임으로 자리매김하기 시작했음을 암시하였다. ‘자강력’ 혹은 ‘자강’이란 표현 역시 마찬가지이다. 2016년에 들어 처음 등장하였던 ‘자강’의 개념은 2017년 신년사에서는 언급 횟수가 5회로 대폭 증가했다. ‘자강’에 대한 강조는 주체사상의 일부인 ‘경제에서의 자립’을 연장한 것이라고 볼 수 있지만, 북한이 현재 처한 대외관계에서도 연유된 것으로 보인다. ‘자강’의 개념은 2016년 신년사에서 처음 등장 후 북한은 4차 핵실험, 그리고 이에 이어 6차 장거리로켓 발사 실험을 감행하였다. 결국 이는 UN안전보장이사회(안보리) 결의 2270호에 입각한 대북 제재로 이어

졌다. 2016년 9월 9일 5차 핵실험의 결과 더 격상된 제재를 의미하는 UN 안보리 결의 2321호가 채택되었다. 이러한 국제 제재 속에서 정권·체제의 생존을 도모하기 위해서는 두 가지 방법을 생각할 수 있다. 하나는 내부적으로 역량에 집중하는 것이고, 다른 하나는 중국 등 주변국들의 도움에 의존하는 것이다. 그러나 주변국에 대한 지나친 의존은 오히려 경제, 외교적 종속을 불러오고, 주변국들이 반드시 이에 호응한다는 보장도 없다. 이러한 상황하에서 독자적인 생존을 모색하기 위해서는 내부 역량의 동원이나 결집을 모색할 수밖에 없으며, 이것이 결국 ‘자강’의 강조로 이어졌다.

둘째는 2017년 신년사에 반영된 ‘수소탄’ 실험, 다양한 공격 수단(탄도미사일)들의 시험 발사 등을 통해 “사회주의강국 건설 위업을 승리적으로 전진시켜나갈 수 있는 위력한 군사적 담보가 마련”되었다고 평가하였다. 이는 이미 ‘핵 강국’ 반열에 올라섰음을 대내적으로 선포하는 동시에 이제는 북한을 사실상 ‘핵보유국’으로 인정하라는 대외적인 메시지를 전달하려는 시도로 볼 수 있다. ‘핵무력’의 강화를 위한 의지는 도처에서 나타났다. “우리식의 위력한 주체무기들을 더 많이 개발·생산하여 선군혁명의 병기창을 억척같이 다져야 합니다”라고 언급한 부분이나, “...핵무력을 중추로 하는 자위적국방력과 선제공격능력을 계속 강화해나갈 것입니다”란 역설은 이의 대표적인 사례라 할 수 있다.

그러나 2017년(평화조성 이전기)과 2018년(평화조성기)의 신년사를 비교해보면 ‘우리’, ‘평화’, ‘경제’의 단어수가 증가하였음을 볼 수 있다. 이는 신년사에서 군사적 긴장상태 완화와 평화적 환경 조성 마련을 강조하며, 이를 위해 3가지 제안 ① 평창올림픽 대표단 파견 및 당국회담, ② 다방면의 남북 교류협력 실현, ③ 남북관계 개선 돌파구마련을 위한 대화 재개 등을 제안 한 부분과 일치한다. 더욱이 2019년 신년사에서 2018년에 비해 더욱 증가한 것을 볼 수 있다. 이는 한반도의 긴장 완화와 평화를 조성하겠다는 의지가 담겨있는 것으로 볼 수 있다. 반면에 ‘핵’과 관련한 단어도 2017년에 비해 2018년에 확연하게 증가한 것을 볼 수 있다. 이는 미국이 북한을 압박하고 공격 할 경우 핵 공격한다는 도발적 노선 유지와 대외적으로 평화를 강조하면서 국제사회의 고립에서 탈피하기 위한 북한의 이중적 메시지가 담겨있음을 알 수 있다. 이는 핵무력 완성 이후의 전략적 지위를 재차 주장하며 ‘평화협정-핵군축’ 주장을 보다 강화하기 위한 사전 정지작업 또는 포석으로도 볼 수 있으며, 2018년 신년사는 핵, 미사일 고도화 의지를 일성수준 보이면서 일련의 유화적 행보를 강조함으로써 대북 제재, 압박 국면을 관리하는데 전술적 목표가 있음을 보여주고 있다고 할 수 있다.

그러나 지난 6년간의 김정은 위원장의 신년사와 비교해도 확연하게 달라진 북한의 ‘평화공세’이며 내용적으로는 매우 과격적이고 과감한 유화 모드를 보여 준 것으로 북한의 정권수립 70주년과 남한 동계올림픽을 거론, “북과 남이 다 같이 의의 있는 해”로 언급, 대표단 파견 및 대화 용의를 표명한 것이다. 또한 ‘우리’를 강조하는 것은 2013년, 2014년, 2016년 신년사에서 외세 공조에 대한 비난이 높았지만, 2018년에는 남북 문제에 대한 남북대화 필요성 강조와 미국에 대한 남북한 공조를 강조 하는 것이다.

2019년 신년사(평화 조성기)에서는 첫째, ‘자력갱생’ 구호 하에 ‘자립경제’ 건설에 자원과 물자 동원, 군수공업 포함 민생경제 기어 주문, 간부 부정부패 근절 강조 등 ‘자립경제’ 용어 7회 등장과 제재 영향에 대해서는 구체적 언급 없이 ‘가혹한 경제 봉쇄와 제재 속에서 비약적 발전’만을 주장하였다. 둘째는 지난해 성과에 ‘대단히 만족’ 평가, 남북합의사항 철저히 이행 및 ‘평화·번영·통일’의 새로운 전성기를 열

자고 적극 의사 표명을 하면서 한반도 항구적 평화지대화, 협력·교류 전면적 확대발전, 전 민족 합의에 기초한 평화적 통일방안 적극 모색 등 제시 하였다. 셋째는 ‘완전한 비핵화’를 재확인 하고 핵무기를 만들지도, 시험하지도 않으며, 사용, 전파하지 않을 것임을(핵무기 4不) 선포하고, 미국측에 ‘상응 행동’을 강조하면서 북미 관계개선에 적극 의사 표시를 하였다.

하노이 북미 정상회담 결렬이후 2019년 4월 15일(평화 조성 침묵기) 북한의 최고인민회의에서 김정은의 시정연설에서 키워드는 ‘우리’, ‘인민’, ‘사회주의’, ‘국가’, ‘공화국’, ‘미국’, ‘건설’, ‘조미’, ‘혁명’, ‘북남관계’, ‘나라’, ‘김일성’, ‘힘’, ‘평화’, ‘자주’와 같은 단어들 이 많이 등장하였다. 먼저 지난 몇 년 동안 신년사에서 등장하지 않은 ‘김일성’ 단어의 등장은 김정은 시대의 새로운 통치이념 제시보다는 김일성·김정일주의의 지속 의지를 보여주는 것으로 김일성·김정일주의의 본질을 인민대중제일주의로 정식화하려는 것이었다. 둘째는 북남관계, 조미, 평화의 단어가 등장하고 있는 것은 연설문에서 “북남관계를 지속적이며 공고한 화해협력관계로 전환시키고, 평화롭고 공동 번영하는 새로운 민족사를 써나가려는 것은 나의 확고부동한 결심”, “계속 진지하고 인내성 있는 노력” 등의 내용이 담긴 것으로 보아 북한이 아직은 인내를 가지고 남·북·미 관계를 풀어가려는 의지를 보여 준 것이라고 판단한다. 연설문에서 트럼프 대통령과의 신뢰관계를 확인하고 있으며, 3차 북미정상회담 개최 필요성을 공감하고 있다. 2019년 말까지 상황악화 조치 자제 및 협상의 여지 개방과 함께 “쌍방이 서로의 일방적인 요구조건들을 내려놓고 각자의 이해관계에 부합되는 건설적인 해법”을 강조하면서 협상안 조정 가능성 시사와 아울러 “우리와 공유할 수 있는 방법론을 찾는 조건에서 정상회담 개최”하자는 내용과 함께 실무회담의 중요성을 암시하고 있었다.

북한 지도자의 신년사와 연설문이 시대적 과제 및 지향점의 요약정보를 담고 있다는 점을 감안했을 때, 국정 운영의 관심 정도가 시대에 따라 신년사와 연설문에 담겨 변화하고 있다는 것을 보여주었다. 이와 같이 시간에 따른 주제의어 군집 결과는 국가적 의제 설정의 추이를 알아볼 수 있는 매우 흥미로운 자료로 사용될 수 있을 것이다.

5. 결 론

본 연구에서 2017년부터 2019년 4월까지 북한지도자의 신년사와 연설문을 텍스트마이닝 방법론을 통해 분석하였다. 북한 인민에 대한 최고지도자의 교시이자 미국과 한국을 비롯한 대외세력에 대한 선언문의 성격을 지닌 북한 신년사와 연설문을 특징지점을 구분하여 텍스트의 특성을 고려하여 북한 신년사 및 연설문 분석에 가장 적합한 분석방법을 고안해서 적용하였다. 구체적으로 신년사 문서의 총체적 구조와 단어 간 상관성을 분석하였다. 그리고 이러한 단어들이 2017년 이전(평화 조성 이전기)과 2018년, 2019년 신년사(평화 조성기), 그리고 2019년 하노이 북미회담 결렬 이후(평화 조성 침묵기)로 나누어 분석함으로써 신년사에 담겨있는 단어의 빈도와 근접성이 북한의 대내외적인 정세와 연관이 있음을 알 수 있게 되었다. 또한 이러한 결과를 텍스트마이닝 기법을 적용하여 도출하고 주요 키워드들을 효과적으로 시각화함으로써 텍스트의 객관적, 정량적 트렌드 분석을 위한 연구결과를 도출 하였다. 본 연구에서 주요 키워드로 선택한 '우리', '경제', '평화', '핵', '통일'로 단어들이 북한의 대남, 조미, 그리고 핵무기관련 정책기조의 변화와 매우 밀접한 상관성을 보여주고 있었다. 또한 이들 단어의 등장위치를 시대적으로 추적한 결과 역시 신년사에서 해당 단어가 표상하는 정책적 목표(조미관계, 대남관계, 그리고 비핵)가 북한정권에게 갖는 중요성이 변화함에 따라 단어의 등장 순서와 위치, 그리고 빈도가 거의 유사하게 변화하고 있음을 알 수 있게 되었다.

본 연구는 북한 지도자의 신년사와 연설문이라는 특정 영역의 문서를 대상으로 분석을 수행하였지만, 제시한 분석 절차 및 기법은 다른 분야에도 충분히 적용이 가능하다. 또한 본 연구에서 북한 관련 신문 기사와 시계열적 데이터를 기반으로 신년사와 연설문을 군집화하고 시각화 분석을 하는 등 북한연구에 다양한 텍스트마이닝 방법론을 도입하고 적용하는 시도들을 할 것이며 본 연구는 그러한 시도의 출발점이라고 볼 수 있다. 텍스트마이닝 분석기법 등 새로운 기술 혁신적 방법으로 북한 연구에 중요한 기여를 할 수 있기를 기대해 본다.

References

- [1] Aggarwal, C. C. and Zhai, C., Mining text data, Springer, 2012.
- [2] Chakraborty, G., Pagolu, M., and Garla, S., Text Mining and Analysis : Practical Methods, Examples, and Case Studies Using SAS, SAS Institute, 2013.
- [3] Chang, J. Y., "A Study on Research Trends of Graph-Based Text Representations for Text Mining", *Journal of the Korea Academia-Industrial Cooperation Society*, Vol. 12, No. 8, 2013, pp. 3677-3688.
- [4] Chen, Y. T. and Chen, M. C., "Using chi-square statistics to measure similarities for text categorization", *Expert Systems with Applications*, Vol. 38, 2011, pp. 3085-3090.
- [5] Cho, G. H., Lim, S. Y., and Hur, S., "An Analysis of the Research Methodologies and Techniques in the Industrial Engineering Using Text Mining", *Journal of the Korean Institute of Industrial Engineers*, Vol. 40, No. 1, 2014, pp. 52-59.
- [6] Cho, S. G. and Kim, S. B., "Finding Meaningful Pattern of Key Words in IIE Transactions Using Text Mining", *Journal of the Korean Institute of Industrial Engineers*, Vol. 38, No. 1, 2012, pp. 67-73.
- [7] Cho, S. G., Cho, J., and Kim, S. B., "Discovering Meaningful Trends in the Inaugural Addresses of United States Presidents Via Text Mining", *Journal of the Korean Institute of Industrial Engineers*, Vol. 41, No. 5, 2015, pp. 453-460.
- [8] Fan, W., Wallace, L., Rich, S., and Zhang, Z., "Tapping the power of text mining", *Communications of the ACM*, Vol. 49, No. 9, 2006, pp. 76-82.
- [9] Go, G. S., Jung, W. K., Shin, Y. G., Park, S. S., and Jang, D. S., "A Study on Development of Patent Information Retrieval Using Textmining", *Journal of the Korea Academia-Industrial Cooperation Society*, Vol. 12, No. 8, 2011, pp. 3677-3688.

[1] Aggarwal, C. C. and Zhai, C., Mining

- [10] Hahm, Y. and Lee, S., "The Distinctiveness of Big Data Business Model in Its Components : A Comparative Analysis of Korea-U.S Cosmetic Big Data Business Cases", *Information Technology and Architecture*, Vol. 13, No. 1, 2016, pp. 63-75.
- [11] <http://www.dailybizon.com/news/articleView.html?idxno=13108>(Search Date : 2019. 4. 20.)
- [12] <https://thaeyongho.com/2019/01/02/김정은-신년사-평가-2-북한에서-신년사가-가진-의미/>(Search Date : 2019. 4. 20.)
- [13] Hu, X. and Liu, H., *Text analytics in social media, Mining text data*, Springer Link, 2012.
- [14] Hung, J. L. and Zhang, K., "Examining mobile learning trends 2003~2008 : A categorical meta-trend analysis using text mining techniques", *Journal of Computing in Higher Education*, Vol. 24, No. 1, 2012, pp 1-17.
- [15] Hymans, J. E. C., "Assessing North Korean Nuclear Intentions and Capacities : A New Approach", *Journal of East Asian Studies*, Vol. 8, 2008, pp. 259-292.
- [16] Judita, P., Stevenson, M., and Gaizauskas, R., "Exploring relation types for literature-based discovery", *Journal of the American Medical Informatics Association*, Vol. 2, 2015, pp. 987-992.
- [17] Kim, H. Y., "Analysis of an Inaugural Address of Korean Presidents Based on Network", *Korea Content Association*, Vol. 3, No. 2, 2013, pp. 67-68.
- [18] Kim, H. Y., Kim, H. G., and Kang, B. M., "A Trend Analysis of Cultural consumption Based on Newspaper Texts", *Journal of KIIS E : Software and Applications*, Vol. 39, No. 3, 2012, pp. 244-251.
- [19] Kim, H., "A Study on Presidential Leadership and Policy Agenda Setting Pattern : A Content Analysis of Korean Presidential Addresses", *Journal of Korean Politics*, Vol. 23, No. 2, 2014, pp. 77-102.
- [20] Kim, J. S., Kim, M. W., and Hyun, B. H., "A Study on Analysis of Patent Information Based Biotechnology Research Trend and Promising Research Themes", *The Korea Society for Innovation Management and Economics*, Vol. 21, No. 2, 2013, pp. 25-56.
- [21] Kim, M. and Koo, P., "A Study on Big Data Based Investment Strategy Using Internet Search Trends", *Journal of the Korean Operations Research and Management Science Society*, Vol. 38, No. 4, 2013, pp. 53-64.
- [22] Kim, M., Notkin, D., Grossman, D., and Wilson, G., "Identifying and summarizing systematic code changes via rule inference", *IEEE Transactions on Software Engineering*, Vol 39, 2013, pp. 45-62.
- [23] Kim, S., Cho, H., and Kang, J., "The Status of Using Text Mining in Academic Research and Analysis Methods", *Journal of Information Technology and Architecture*, Vol. 13, No. 2, 2016, pp. 317-329.
- [24] Kim, Y., Kim, N., and Jeong, S., "Stock-Index Invest Model Using News Big Data Opinion Mining", *Journal of Intelligence and Information Systems*, Vol. 18, No. 2, 2012, pp. 143-156.
- [25] Kim, Y., Tian, Y., Jeong, Y., Jihee, R., and Myaeng, S. H., "Automatic discovery of technology trends from patent text", *Proceedings of the ACM symposium on Applied Computing*, 2009, pp. 1480-1487.
- [26] Lee, C. and Moon, H., "Study on analysis of North Korea's news trends associated with provocations using text mining", *Journal of National Defense Studies*,

- Vol. 59, No. 4, 2016, pp. 103-124.
- [27] Lee, M. and Kim, H., "Construction of Event Networks from Large News Data Using Text Mining Techniques", *Journal of Intelligence Information System*, Vol. 24, No. 1, 2018, pp. 183-203.
- [28] Lee, Y. J., Seo, J. H., and Choi, J. T., "Fashion Trend Marketing Prediction Analysis Based on Opinion Mining Applying SNS Text Contents", *The Journal of Korean Institute of Information Technology*, Vol. 12, No. 12, 2014, pp. 163-170.
- [29] Lim, E. T., "Five trends in presidential rhetoric : An analysis of rhetoric from George Washington to Bill Clinton", *Presidential Studies Quarterly*, Vol. 32, No. 2, 2002, pp. 328-348.
- [30] Liu, B., "Sentiment analysis and opinion mining", *Synthesis Lectures on Human Language Technologies*, Vol. 5, No. 1, 2012, pp. 1-167.
- [31] Liu, B., "Sentiment Analysis and Subjectivity", *Handbook of Natural Language processing*, Vol. 2, 2010, pp. 627-666.
- [32] Myung, J., Lee, D., and Lee, S., "A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary", *Journal of KISS : Software and Applications*, Vol. 35, No. 6, 2008, pp. 392-403.
- [33] Narayanan, R., Liu, B., and Choudhary, A., "Sentiment Analysis of Conditional Sentences", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Vol. 1, 2009, pp. 24-31.
- [34] Oh, G. S. and Lee, K. H., Analysis of the Kim Jong-un Regime's South Korea Policy and Unification Discourse through Text Mining, Korea Institute for National Unification, 2016.
- [35] Pai, M. Y., Chen, M. Y., Chu, H. C., and Chen, Y. M., "Development of a semantic-based content mapping mechanism for information retrieval", *Expert Systems with Applications*, Vol. 40, 2013, pp. 2447-2461.
- [36] Park, H., Seo, W., Coh, B., Lee, J., and Yoon, J., "Technology Opportunity Discovery Based on Firms' Technologies and Products", *Journal of the Korean Institute of Industrial Engineers*, Vol. 40, No. 5, 2014, pp. 442-450.
- [37] Park, J. H., Park, E., and Jo, D., "Automated Text Analysis of North Korean New Year Addresses, 1946~2015", *Korean Political Science Review*, Vol. 49, No. 2, 2015, pp. 27-62.
- [38] Rich, T. S., "Deciphering North Korea's Nuclear Rhetoric : An Automated Content Analysis of KCNA News", *Asian Affairs*, Vol. 39, 2012, pp. 73-89.
- [39] Ryu, J., Han, C., and Shin, H., "Sector Investment Strategies Using Big Data Trends", *Information Technology and Architecture*, Vol. 13, No. 1, 2016, pp. 111-121.

■ 저자소개



Chul-Soo Park

Chul-Soo Park is Working as a Professor at the Department of Business Administration of Halla University. He is also the director of the Northeast Asia Economic Research Institute. He received his M.S., Ph.D. from the Business College at KAIST. His current research interests include North-Korea Business and Economy, Entrepreneurship Management, Big Data Analysis in Accounting and Finance, AI Application in Management, Intellectual Capital, and etc.