

북한 스케치

2000년대 신문 빅 데이터에 표상된 북한의 모습*

김일환(고려대학교), 이도길(고려대학교), 강진웅(고려대학교)**

이 연구는 2000년대 최근 14년 동안(2000~2013년)의 주요 일간지를 대상으로 ‘북한’과 관련한 단어 사용 양상을 계량적으로 관찰함으로써 한국 언론에 보도된 북한의 모습을 객관적으로 포착해 보는 데 그 목적이 있다. 북한에 대한 우리 사회의 높은 관심에도 불구하고 북한이 한국 언론에 표상된 모습을 계량적, 정량적으로 파악해 보려는 시도는 지금까지 매우 제한된 양상으로 이루어져 왔다. 이 연구에서는 14년 치의 신문 빅 데이터인 ‘물결21’ 코퍼스를 기반으로 하여 북한의 공기어와 의미운율을 분석함으로써 한국 언론에 표상된 북한의 모습을 제시해 보았다. 특히 북한의 공기어를 국제 정세의 변화, 남남갈등의 변화 추이, 북한의 생활상, 남북 지원 사업의 소멸, 남북갈등의 심화, 북핵 문제, 탈북 문제 등으로 분류하여 그 관련성의 변화를 살펴보았다. 또한 북한 관련 공기어의 의미운율을 긍정성과 부정성으로 분류하여 분석해 봄으로써 북한과 관련한 우리의 감정이 어떻게 변화하였는지를 거시적으로 포착해 보려 시도하였다. 이를 통해 2000년대 들어 북한과 관련한 한국 언론, 나아가 우리 사회의 인식이 어떠한 양상으로 전개되어 왔는지를 거시적으로 규명할 수 있었다.

주제어: 북한, 신문 빅 데이터, ‘물결21’ 코퍼스, 계량적 접근, 통계적 분석, 공기어, 의미운율, 언어 사용 양상

* 이 연구는 2007년 정부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2007-361-AL0013).

** 김일환(제1저자), 이도길(공동저자), 강진웅(교신저자).

1. 서론

이 연구는 2000년대 최근 14년 동안(2000년~2013년)의 주요 일간지를 대상으로 ‘북한’과 관련한 단어 사용 양상을 계량적으로 관찰함으로써 한국 언론에 보도된 북한의 모습을 객관적으로 포착해 보는 데 그 목적이 있다.

지금까지 한국 언론에 보도된 북한의 모습은 어떠했는가?

북한에 대한 우리 사회의 높은 관심에도 불구하고 우리는 북한에 대한 정보를 언론과 방송에 전적으로 의존하고 있다고 해도 과언이 아니다. 특히 주요 언론이 보도하는 북한의 모습은 우리 사회에 미치는 영향력이 막대할 뿐 아니라 우리가 접하는 북한 관련 정보의 대다수를 차지하고 있다는 점에서 언론 속에서 북한이 어떻게 다루어지고 있는지를 면밀히 살펴보는 것은 매우 중요하다.

특히 북한은 끊임없이 한국 사회에 대한 위협적인 행동을 취하기도 하고 때로는 남북 정상 회담과 이산가족 상봉 등에서와 같이 친밀한 한민족의 모습을 보여주기도 하는 등 우리에게 매우 다양한 모습으로 표상되어 왔다. 이 연구에서는 『조선일보』, 『동아일보』, 『중앙일보』, 『한겨레신문』(발간순)의 2000년대 신문 기사 전체로 구성된 ‘물결21’ 코퍼스라는 신문 빅 데이터를 기반으로 해서 ‘북한’에 대한 언론의 주요 이슈들을 계량적인 방법론을 활용하여 분석해 보고자 한다.¹⁾ 신문 빅 데이터에 기반을 둔 계량적 접근을 통해 다소 모호하거나 추상적으로 알고 있던 북한의 모습이 좀 더 명시적이고 객관적인 양상으로 드러날 수 있을 것이다. 특히 이러한 방법은 연구자나 기존

1) ‘물결21’ 코퍼스에 대한 자세한 내용은 김일환 외(2013)에서 상세히 밝히고 있다.

연구의 선입견이나 주관을 배제할 수 있다는 점에서 북한의 모습을 좀 더 객관적으로 관찰하는 데 기여할 수 있을 것으로 기대된다.²⁾

본고에서 적용하고 있는 계량적 분석 방법은 이미 최근의 여러 논 의들에서 활용된 바 있다. 특히 Baker 외(2012)에서는 ‘muslim’이라는 단어가 영국의 주요 일간지에 어떻게 표상되어 있는지를 2000년대 신문을 기반으로 분석한 바 있는데, 여기서는 주로 연어 관계와 의미 운율을³⁾ 관찰함으로써 2001년에 미국에서 일어난 9/11과 2005년의 런던 지하철 테러 사건을 계기로 무슬림에 대한 영국 사회의 인식이 어떻게 변화했는지를 계량적으로 규명한 바 있다.

이에 비해 국내에서는 송정호·이상원(2015), 송태민(2015) 등에서와 같이 단편적으로 빅 데이터를 이용하여 북한 관련 문제를 분석해 보 려는 시도가 나타나기 시작하고 있으나 본격적인 거시적 차원의 논의 는 이루어지지 못하고 있다.

이 논문의 구성은 다음과 같다. 2절에서는 분석 대상 자료의 규모 와 특성, 그리고 연구 방법에 대한 논의를 진행하고, 3절에서는 ‘북한’ 의 공기어에 대한 계량적 분석을 수행한다. 그리고 4절에서는 ‘북한’ 의 공기어에 대해 의미운율 분석을 시도하고 마지막으로 5절에서는

2) 물론 이러한 논의는 한국의 언론에 보도된 내용에 국한된 것이다. 익명의 심사자 께서도 지적한 바와 같이 한국 언론에서 보도한 내용이 북한의 실상을 얼마나 사실대로 보도하느냐 하는 점이 중요한 요건임에는 틀림없으나 이 연구에서는 언론에 표상된 북한의 모습을 규명하는 데 초점이 있을 뿐 언론 보도 내용의 사실성 여부를 논하는 것은 본고의 범위를 벗어난다. 또한 여기서 ‘명시적이고 객관적인 양상’이라는 것은 적어도 언어 사용 양상이라는 구체적인 결과에 기인 하는 것일 뿐 사실 관계의 추론에서 도출된 내용이 아님을 밝혀둔다.

3) 연어(連語, collocation)는 임의의 두 단어가 연속해서 출현하는 단어를 지칭하는 것이고 의미운율은 단어들이 함께 출현하면서 생기는 새로운 의미를 나타내는 개념이다. 이에 대한 자세한 논의는 4절을 참조할 것.

본론의 내용을 요약하고 향후 과제를 제시할 것이다.

2. 대상 자료와 방법

1) 대상 자료의 규모와 특성

이 연구에서 연구 대상으로 삼고 있는 ‘물결21’ 코퍼스는 전체 어절 규모가 약 5.9억 어절에 이르는 대규모의 언어 자원, 빅 데이터라 할 수 있다. <표 1>은 ‘물결21’ 코퍼스의 연도별 규모를 정

<표 1> ‘물결21’ 코퍼스의 연도별 규모

연도	기사 수	어절 수
2000년	231,361	40,377,975
2001년	220,109	39,728,480
2002년	214,941	42,298,083
2003년	205,635	42,959,317
2004년	200,260	40,652,017
2005년	180,413	38,062,190
2006년	162,055	37,539,133
2007년	177,743	42,432,005
2008년	176,697	42,322,623
2009년	164,314	43,230,845
2010년	170,338	48,307,337
2011년	157,755	44,425,114
2012년	153,227	44,017,316
2013년	152,348	45,367,321
전체	2,567,196	591,719,756

리한 것이다.⁴⁾

‘물결21’ 코퍼스가 4개의 일간지로 구성되어 있다는 점을 감안하면 신문 기사는 신문사마다 매년 1천만 어절의 텍스트를 생산하는 것으로 짐작할 수 있다(물론 신문사 간 차이는 무시할 경우). 이러한 규모의 텍스트 자원은 최근 빅 데이터 연구에서 주목받고 있는 소셜 네트워크서비스(SNS) 자료에 비해 규모면에서는 작을 수 있으나 신문 텍스트리는 공공적 특성을 고려할 때 사회, 문화 연구를 위한 기반 자료로서는 손색이 없다. 특히 신문은 SNS가 등장하기 훨씬 오래전부터 존재해 왔다는 점에서 거시적인 차원의 계량적 연구 자료로서 그 가치가 크다고 할 수 있다.

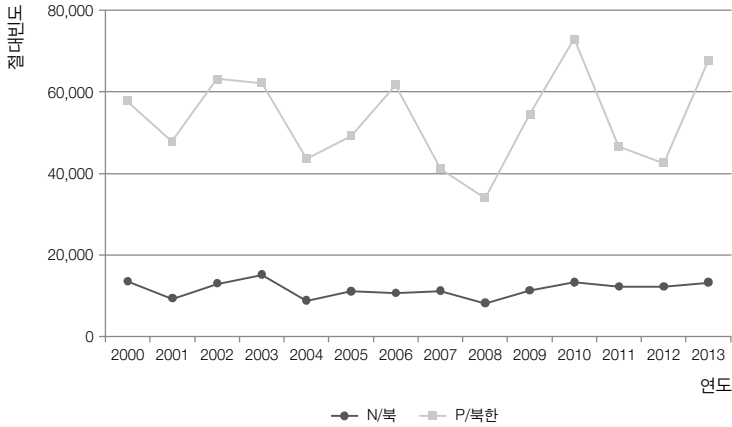
2) 연구 방법

이 연구에서는 신문 빅 데이터에 나타난 단어의 사용 양상에 주목한다.⁵⁾ 이를 위해 먼저 고려되어야 할 문제는 복한을 비롯한 관련

4) 익명의 심사자께서는 ‘물결21’ 코퍼스에 포함된 신문의 균형성에 문제를 제기하기도 하였다. 즉, 보수적인 성향의 신문이 셋인데 비해 진보적인 성향의 신문은 『한겨레신문』뿐이므로 이들을 뭉뚱그려 다루면 신문사 간의 보도 차이를 포착할 수 없다는 것이다. 이러한 의견은 대체로 동의하지만 단어 사용의 전반적인 양상을 살피고 이를 통해 복한 관련 논의의 흐름을 파악하는 데에는 큰 문제가 없다고 생각한다. 특히 이 논문에서는 신문사 간의 복한 관련 보도 차이를 밝히는 것에는 관심이 없고 언론에 표상된 복한의 전반적인 모습을 기술하는 데에 목적이 있기 때문에 이러한 문제는 심각하지 않다고 본다.

5) 익명의 심사자께서는 본 논문의 연구방법에는 기본적으로 동의할 수 있으나 구체적인 문제제기, 연구문제가 명확하게 드러나지 않고 있다는 지적을 한 바 있다. 사회과학적인 관점에서 가설을 설정하고 이에 대한 근거를 밝히는 과정에서 연구문제의 설정은 가장 중요한 핵심임에 두말할 나위가 없다. 그러나 이 연구는 언론에 표상된 대규모의 복한 관련 기사를 분석함으로써 복한과 관련한 이슈와 사회적 관심이 어떻게 변화해 왔는지를 기술하는 데에 목적이 있다. 즉,

<그림 1> ‘북한’과 ‘북’의 연도별 사용 빈도



단어들의 사용 빈도이다. <그림 1>은 ‘북한’과 ‘북’의 연도별 사용 빈도를 제시한 것으로서 단어의 사용 빈도는 해당 시기의 관심사를 반영해 주는 척도로 해석할 수 있다.⁶⁾ 북한의 빈도는 연도별로 평균 약 6만 회 정도에 육박할 정도로 높은 빈도를 보이는데⁷⁾ 이는 전체 고유명사 빈도 순위 가운데 5위에 해당한다. 2000년의 고유명사 빈도를 기준으로 한다면 ‘북한’은 미국, 한국, 서울, 일본에 이은 5위에

특정한 연구문제를 선형적으로 가정하는 방식이 아닌, 데이터가 보여주는 것을 기술하는 과정에서 연구문제를 도출하는 데이터-유도(data-driven) 접근 방식을 취하고 있다는 점을 상기할 필요가 있다.

- 6) 본고에서 사용하는 단어 사용 빈도나 공기어 정보 등은 ‘물결21’ 코퍼스를 활용하기 위해 개발된 ‘웹 기반 코퍼스 분석 도구’(corpus.korea.ac.kr)를 참조할 수 있다. 이에 대한 자세한 활용 방법은 이도길·최재용(2015)을 참고할 것.
- 7) 물론 이 수치는 평균에 기댄 논의일 뿐이다. 실제 연도별 북한의 빈도를 살펴보면 2008년과 2012년에는 다른 해에 비해 상대적으로 빈도가 낮다. 이에 대한 원인은 매우 다양한 요인들과 관련되어 있을 것인데, 현재로서는 이를 명확하게 밝히는 데 한계가 있다. 이에 대한 상론은 후고를 기약한다.

해당하며 이는 ‘북한’이 한국 언론에서 차지하는 높은 비중을 간접적으로 보여주고 있다.⁸⁾

한편 이 연구에서는 공기어(co-occurrence word) 정보를 적극적으로 분석에 활용할 것이다. 공기어는 김일환 외(2010) 등에서 이미 논의된 바와 같이 특정한 범위 안에서 대상어와 함께 출현하는 단어들로 규정될 수 있으며, 연속해서 출현할 필요는 없다.

이 연구에서는 ‘문장(sentence)’을 단위로 ‘북한’의 공기어를 추출하고 그 특징을 연도별로 분석하는 방식을 취하였다. 이때 우연히 출현한 공기어를 배제하기 위한 방법으로 t-점수를 활용하였다.⁹⁾ 이를 통해 ‘북한’과 관련성이 높은 공기어들, 그리고 관련성에 변화가 나타나는 공기어들을 계량적으로 분석해 낼 수 있으며 이를 통해서 북한과 관련한 언어 사용, 나아가 언론에 비친 북한의 모습을 유추해 볼 수 있다.

다음의 <표 2>는 2000년부터 2013년까지 보도된 신문 기사 중 ‘북한’의 공기어를 추출한 것이다. 여기서 $f(W1)$ 는 대상어인 ‘북한’의 전체 빈도를, $W2$ 는 북한과 같은 문장에 출현한 명사를 나타낸다 (N 은 일반명사, P 는 고유명사). 또 $f(W2)$ 는 공기어의 출현빈도를, $F(W1W2)$ 는 대상어와 공기어가 한 문장에 출현한 빈도를 나타내며 마지막으로 t-score는 대상어와 공기어가 함께 출현한 정도를 표시한

8) 여기서는 ‘절대빈도’를 기준으로 논의하였다. 연도별로 신문의 규모가 조금씩 다르므로 좀 더 정확한 비교를 위해서는 상대빈도를 참조하는 것이 바람직하겠으나 2000년대 신문의 경우 절대빈도와 상대빈도의 차이가 크지 않으므로 여기서는 절대빈도를 기준으로 하였다. ‘물결21’의 웹 기반 분석 도구에서는 상대빈도뿐 아니라 월별 사용빈도까지 확인할 수 있다.

9) 공기어의 t-점수 측정 방법에 대한 상세한 논의는 김일환·이도길·강범모(2010) 등을 참조할 것.

< 표 2 > '북한'의 주요 공기어 상위 50개(t-점수 기준)

W1	f(W1)	W2	f(W2)	f(W1W2)	t-score
P/북한	744860	N/핵	192292	120643	336.01
P/북한	744860	P/미국	1435450	124739	270.04
P/북한	744860	N/회담	255234	72852	250.57
P/북한	744860	N/문제	1137330	90230	222.93
P/북한	744860	P/중국	898412	81599	221.32
P/북한	744860	N/미사일	85515	43761	200.83
P/북한	744860	P/미	401590	53505	195.80
P/북한	744860	N/남북	180671	41090	184.47
P/북한	744860	N/인권	128445	38988	184.15
P/북한	744860	N/대북	111386	38182	183.74
P/북한	744860	P/김정일	70704	32829	173.21
P/북한	744860	N/북	161701	33377	164.59
P/북한	744860	N/대화	151463	31954	161.42
P/북한	744860	P/남한	39329	26187	156.85
P/북한	744860	N/정부	1290350	65856	153.77
P/북한	744860	N/실험	96112	26810	151.73
P/북한	744860	N/위원장	314549	34529	151.19
P/북한	744860	N/관계	361096	35792	150.14
P/북한	744860	N/발사	56510	24738	149.93
P/북한	744860	N/국방	80869	23823	143.63
P/북한	744860	N/체제	162902	26237	141.41
P/북한	744860	N/방문	238452	28667	140.50
P/북한	744860	N/도발	27511	20470	139.14
P/북한	744860	N/군	288929	29817	138.45
P/북한	744860	P/한반도	105691	22826	136.77
P/북한	744860	N/핵무기	32425	18727	132.00
P/북한	744860	N/위협	78784	19979	129.95
P/북한	744860	N/제재	57441	18483	127.31
P/북한	744860	N/지원	646412	37847	126.57
P/북한	744860	P/평양	59584	18026	125.18
P/북한	744860	N/당국	127526	19997	122.96
P/북한	744860	N/국제	378800	28234	121.91
P/북한	744860	N/통일	111147	18687	120.07

P/북한	744860	N/합의	202629	21520	118.44
P/북한	744860	P/유엔	95802	17632	118.03
P/북한	744860	N/외교	149135	19342	117.14
P/북한	744860	N/해결	217828	21613	116.70
P/북한	744860	P/부시	142779	18894	116.21
P/북한	744860	N/가능성	396612	27018	115.01
P/북한	744860	N/주민	347728	25112	113.58
P/북한	744860	N/대통령	1075940	46417	113.29
P/북한	744860	N/정상	225131	20830	112.42
P/북한	744860	N/장관	398280	26234	111.67
P/북한	744860	N/평화	157138	18304	111.53
P/북한	744860	N/보도	285180	22307	110.30
P/북한	744860	N/탈북자	44513	13904	110.19
P/북한	744860	N/정책	549651	30410	109.91
P/북한	744860	N/말	2782750	89788	109.67
P/북한	744860	N/관련	725180	35394	109.28
P/북한	744860	N/주장	494508	28197	107.68

것이다.¹⁰⁾

공기어의 관련성은 시기에 따라 변화할 수 있다. 즉, 같은 공기어라고 하더라도 관련성의 정도(t -점수)에서 연도별로 차이가 있는데 이는 시기별 이슈와 관련지어 해석할 수 있다. 이를테면 <그림 2>는 연도별로 ‘북한’과 가장 관련성이 높은 상위 20개의 공기어 목록을 정리한 것이다.

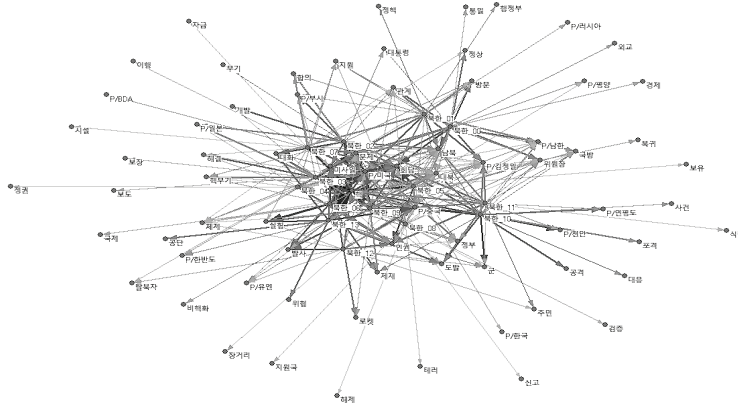
t -점수를 기준으로 정리된 <그림 2>의 명사들을 살펴보는 것만으로도 우리는 2000년대 초반 10여 년 동안 한국 사회의 언론에 표상된 북한 관련 이슈들을 포괄적으로 이해할 수 있게 된다. 또한 연도별 공기어들을 네트워크로 시각화하여 제시하는 것도 북한 문제를 정리

10) t -score 산출을 위한 구체적인 과정은 김일환·이도길·강범모(2010)를 참조할 것.

<그림 2> '북한'의 연도별 공기어 중 관련성이 높은 상위 20개 목록

2000년	2001년	2002년	2003년	2004년	2005년	2006년	2007년	2008년	2009년	2010년	2011년	2012년	2013년
회담	P/미국	원	원	원	원	원	원	원	원	P/중국	P/중국	발사	원
남북	위협장	P/미국	P/미국	P/미국	회담	P/미국	회담	P/미국	P/미국	도발	P/김정일	P/중국	P/중국
방문	P/김정일	문제	문제	문제	P/미국	실험	P/미국	문제	미사일	군	회담	미사일	실험
P/남한	남북	대화	회담	회담	인권	미사일	문제	인권	발사	찬안	인권	인권	도발
P/미국	미사일	P/미	P/미	인권	문제	회담	합의	회담	회담	핵	핵	핵	P/미국
미사일	방문	개발	P/중국	P/중국	P/미	발사	P/미	남북	P/미	회담	도발	로켓	대화
관계	국방	P/부시	해결	남북	P/중국	남북	남북	대북	실험	P/연평도	남북	도발	미사일
문제	대화	미사일	원	P/미	핵무기	문제	원	P/미	대북	공격	주민	P/미국	회담
경상	회담	회담	핵무기	지원	원	제재	대북	위협장	P/중국	P/김정일	위협장	경거리	위협
P/김정일	관계	원	대북	P/김정일	대북	대북	지원	원	대북	대북	대화	군	P/미
P/중국	P/미	남북	대화	해결	P/김정일	인권	P/BDA	관계	P/미국	P/김정일	위협장	경거리	위협
위협장	대북	정부	P/일본	원	북귀	P/미	시설	P/김정일	제재	도격	대북	P/김정일	인권
대통령	문제	대북	미사일	탈북자	남북	정부	미사일	정부	북	정부	찬안	문제	대북
지원	정책	P/일본	개발	P/남한	정부	P/유엔	관계	P/남한	P/남북	P/미국	국방	문제	P/한반도
P/중앙	P/중국	합의	남북	대북	P/남한	핵무기	이행	타러	P/김정일	제재	군	대북	남북
대북	P/부시	관계	인권	체제	원	원	경상	검증	로켓	위협장	정부	주민	군
P/미	P/남한	무기	보장	평문	해결	P/김정일	실험	지원국	정부	남북	P/미국	실험	계제
국방	원	P/중국	체제	P/일본	체제	국제	해결	지원	위협장	대중	P/미	정부	문제
경제	항공부	핵무기	보도	핵무기	위협장	남북	P/김정일	P/중국	P/유엔	P/남한	식량	P/미	발사
원	P/러시아	대통령	P/한반도	정부	보유	P/부시	체제	국방	관계	국방	체제	P/남한	P/유엔

<그림 3> '북한'의 연도별 공기어 네트워크(상위 20개)

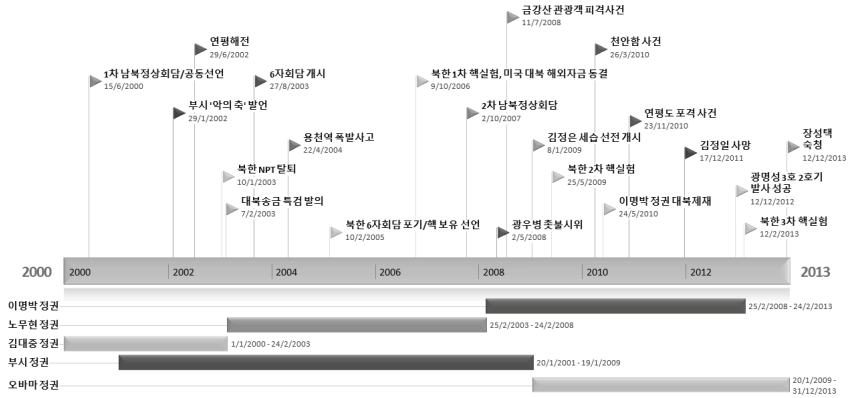


하는 또 다른 방법이다.11)

한편 북한 문제에 대한 기존의 논의를 상기하기 위해 북한 관련 주요 연표를 <그림 4>에서 정리하였다. <그림 4>는 북한의 공기어

11) 그러나 이번 연구에서는 네트워크에 의한 공기어 시각화는 이 정도로만 제시할 것이다. 본문에서는 주로 공기어의 관련성 변화 추이를 분석하는 데에 초점을 둘 것이다.

<그림 4> ‘북한’ 관련 주요 연표



들을 사회, 문화적 배경 속에서 파악하는 데에 도움을 줄 것이다.

마지막으로 이 연구에서는 ‘북한’의 의미운율을 분석할 것이다. 의미운율은 최근 빅 데이터 연구에서 주로 활용되고 있는데 남길임 (2012)에서도 논의된 바와 같이 대상어와 공기어 가운데 일부는 빈번하게 자주 쓰이는 가운데 특정한 의미적 효과를 가져올 수 있으며, 이러한 의미적 효과는 크게 긍정성과 부정성으로 구분할 수 있다. 북한 관련 공기어의 의미운율을 시기별로 관찰, 분석함으로써 ‘북한’에 대한 인식을 좀 더 거시적인 차원에서 파악할 수 있게 될 것이다.

3. ‘북한’의 공기어 추이 분석

이 절에서는 ‘북한’의 공기어들 가운데 서로 주제적으로 관련성이 높은 공기어들을 먼저 분류하고 이들의 변화 추이를 살펴보는 방식으로 논의를 진행할 것이다.

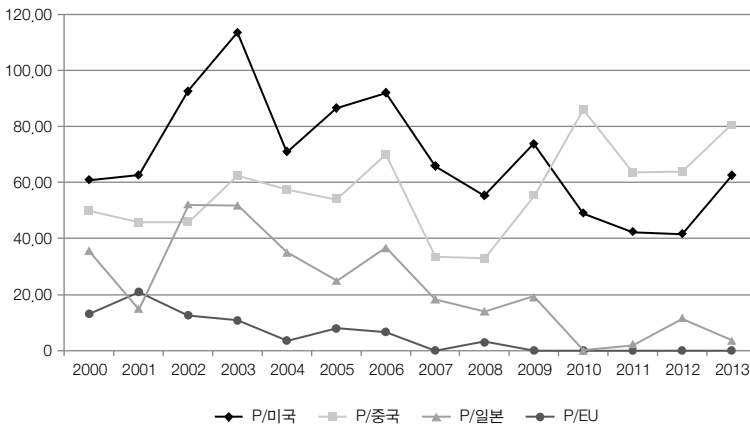
1) 북한을 둘러싼 국제 정세의 변화

북한의 공기어 가운데 국가명의 관련성이 어떻게 변화하는지를 살펴봄으로써 북한을 둘러싼 국외 정세의 변화 양상을 가늠해 볼 수 있다. <그림 5>는 ‘미국, 중국, 일본, EU’의 관련성이 시기별로 어떻게 변화해 왔는지를 보여준다(x축은 연도, y축은 공기어의 관련성 척도인 t점수를 나타냄, 이하 같음).

<그림 5>에서 두드러지는 추이는 ‘중국’의 관련성 증가와 ‘일본, EU’의 관련성 감소로 요약된다. 중국의 관련성 증가는 다음 (1)의 기사에서 확인할 수 있듯이 2010년 무렵 북한의 중국 의존도가 심해진 시점과 대체로 일치한다.

- (1) 중국의 한반도 통일관(統一觀)이 결코 우리에게 우호적이라고 할 수만은 없는 상황에서 북한의 중국 의존도가 높아질수록 한반도 통일에

<그림 5> ‘북한’ 대외 관계 관련 공기어의 관련성 변화



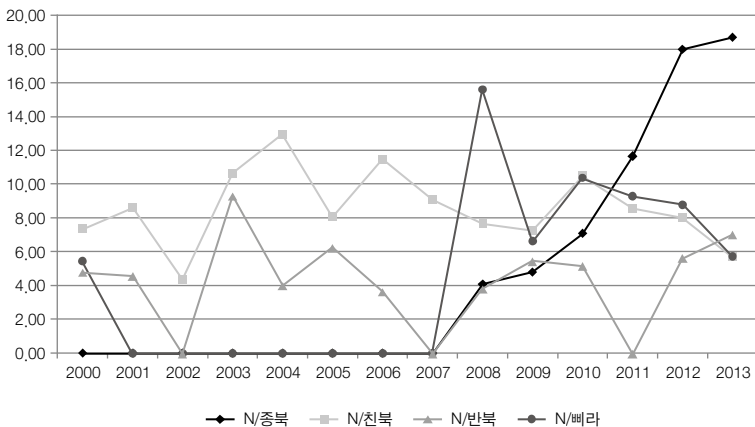
대한 중국의 발언권은 강화되고, 통일 문제의 중국화(中國化)도 가속화 될 것이다(『조선일보』, 2010년 1월 1일).

한편 ‘일본’과 ‘EU’는 2009년 광명성 인공위성 발사 및 핵실험 등의 여파로 관계가 급격히 악화하면서 2010년 이후 관련성이 크게 감소한 것으로 보인다.

2) 남남갈등의 변화 추이

이념적 대립과 관련한 소위 ‘남남갈등’과 관련한 공기어들의 관련성 변화는 역대 정권의 성향과 밀접한 관련을 보이는 것으로 해석된다. <그림 6>에서 알 수 있듯이 MB 정부가 들어선 2008년부터 ‘중북’이 처음으로 높은 관련성을 보인 이래 지속적인 증가 추이를 보이고 있으며 이는 ‘빼라, 전단, 살포’ 등의 공기어와도 연관되어 있다.

<그림 6> ‘남남갈등’ 관련 공기어의 관련성 변화



한편 <그림 6>에서는 제시되지 않았으나 ‘좌파, 뉴라이트’ 등의 관련성 변화와도 함께 해석될 수 있을 것이다(2).

(2) 뉴라이트는 좌파의 ‘뜨거운 감자’인 북한 인권 문제를 제기했고 ‘평화’와 ‘인권’ 중 어느 쪽을 중시할 것인지 입장이 갈렸다(『조선일보』, 2006년 1월 27일).

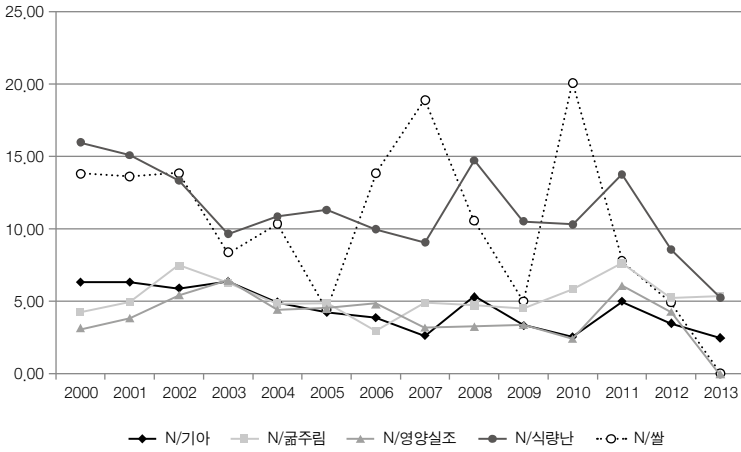
3) 북한의 생활상

북한의 생활상에 대한 정확한 실상을 언론에 나타난 기사만으로 추측하기에는 한계가 있다. 특히 <그림 7>과 <그림 8>에서 보는 바와 같이 북한의 공기어 중 생활상과 관련한 단어들은 그 유형이 매우 제한되어 있다는 점에서 북한의 생활상을 재현하기에는 무리가 있다. 더욱이 북한의 생활상 자체에 대한 관심은 북핵이나 이념 등의 이슈에 비해 무게가 떨어지는 것도 사실이다.

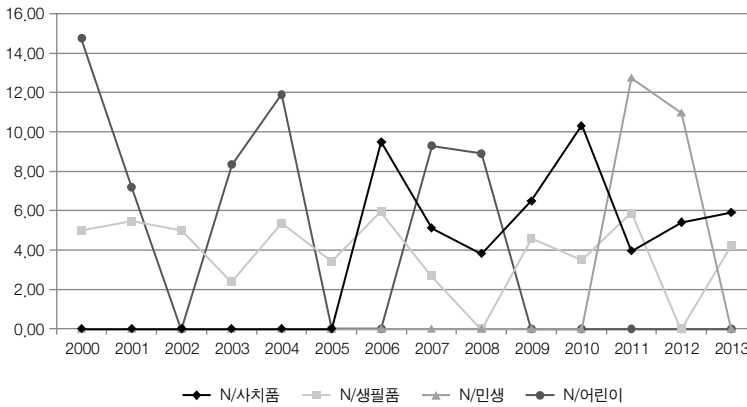
먼저 <그림 7>에서는 ‘기아, 굶주림, 영양실조, 식량난’ 등과 같이 북한의 어려운 실상을 보여주는 공기어들이 2000년대에도 ‘북한’과 높은 관련성을 지속적으로 보이고 있다는 점이 관찰된다. 단 2011년을 기점으로 ‘식량난’의 관련성은 감소 추이를 보인다고 볼 수 있는데 이는 북한의 식량 사정이 다소 나아진 현상과 관련된 것으로 해석할 수 있다. 한편 ‘쌀’은 (3)~(4)의 기사에서 확인할 수 있듯이 북한의 실상과 관련된 것이라기보다는 우리 정부의 대북 지원 정책과 관련된 것으로 2007년과 2010년에 그 관련성이 크게 높아진 점을 발견할 수 있다.

(3) 이재정 통일부 장관은 8일 북한에 차관 형식으로 제공해 온 쌀

<그림 7> 생활 관련 공기어의 변화(1)



<그림 8> 생활 관련 공기어의 변화(2)



지원 방식을 무상 지원으로 전환하는 방안을 검토하고 있다고 밝혔다 (『동아일보』, 2007년 1월 9일).

(4) 정부는 이번 지원을 ‘수해 대처용’으로 규정하고 ‘대규모 쌀 지원’을 하려면 북한의 태도 변화가 필요하다’고 하지만, 이는 남북관계를

주도적으로 풀려는 자세가 아니다(『한겨레신문』, 2010년 9월 14일).

한편 <그림 8>의 공기어들도 일견 북한의 생활상과 관련된 것으로 해석될 수 있으나 그 자체만으로는 구체적인 실상을 파악하기 어려울 뿐 아니라 일부는 생활상과 무관한 정치적인 이슈와 관련된 것이라는 점을 다음 기사를 통해서도 확인할 수 있다.

(5) 워싱턴의 한 소식통은 “사치품 유입 차단은 북한 주민에게 전혀 영향을 미치지 않는다”며 “북한 지도부의 결속력을 약화시키려는 의도가 읽힌다”고 말했다(『동아일보』, 2006년 10월 11일).

(6) 대북 심리전(戰)이라는 이름을 붙이기는 했지만 북한에 전단을 날려 보내고 생필품을 살포하는 것은 북한 주민의 알 권리와 생존권을 보호하는 인권활동이다(『동아일보』, 2011년 2월 28일).

(7) 한나라당 이주영 정책위의장은 31일 “여·야 원내대표가 전날 합의한 ‘북한 민생 인권법’은 상식적으로 말이 안 된다”(『조선일보』, 2011년 6월 1일).

그렇다면 <그림 8>에 있는 단어 중 북한의 생활상과 관련한 단어로는 ‘어린이’ 정도만이 남는다. 이는 국내의 언론들이 북한의 생활상과 관련한 문제를 매우 제한적으로 취급했거나 소홀히 했음을 보여주는 것으로도 해석될 수 있는데, 이를 통해 국내에서 북한의 생활상을 언론을 통해 정확히 추측하기는 어렵다는 사실을 알 수 있다.

4) 남북 지원 사업의 소멸

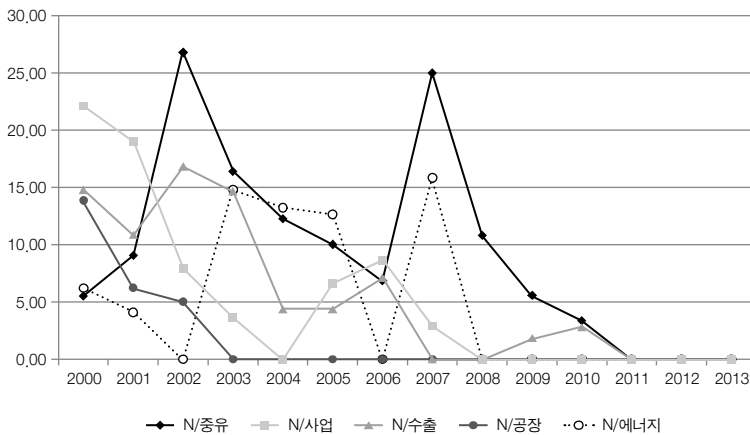
‘햇볕정책’으로 집약되는 “국민의 정부” 시절의 주요 대북 정책에
 는 다양한 남북 지원 사업이 포함되었다. <그림 9>에 나타난 ‘사업’
 은 남북 경협 사업을 포함, 다양한 지원 사업으로 이어지다가 2008년
 이후 관련성이 소멸된다(8~9).

(8) 국내외 기업 경영진들이 대북 사업 규모를 늘리거나 신규 투자를
 탐색하기 위해 다음달 대거 북한을 방문한다(『중앙일보』, 2000년 4월
 13일).

(9) 예를 들어 그동안은 금강산관광·개성공단 사업을 통해 북한에 현
 금이 들어가지 않도록 해야 한다고 해왔다(『조선일보』, 2007년 7월 6일).

한편 2002년과 2007년에 관련성이 급증한 ‘중유’는 (10)~(11)의 기

<그림 9> 남북 지원 사업 관련 공기어의 변화



사에서 보이듯이 미국의 경수로 건설 지원과 관련한 것으로 우리 정부의 대북 정책과 연관된 것은 아니다.

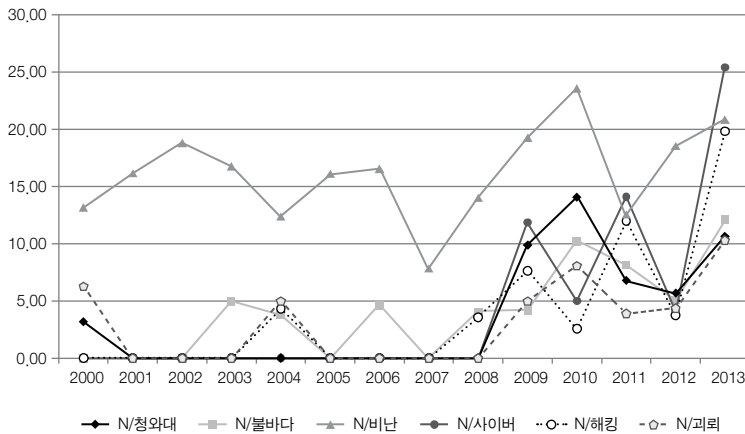
(10) 미국은 94년에 맺은 북-미 기본조약을 통해서 북한이 미사일 실험과 핵시설에 대한 사찰을 받는 조건으로, 중유 지원과 경수로 건설에 협조하기로 했다(『한겨레신문』, 2002년 2월 27일).

(11) 북한은 베이징 6자회담에서 핵 시설 동결의 대가로 중유(重油) 지원 재개를 적극 주장하고 있는 것으로 8일 알려졌다(『조선일보』, 2007년 2월 9일).

5) 남북갈등의 심화

남북관계가 급격히 경색하면서 남북갈등과 관련한 단어의 관련성이 크게 증가하고 있다. <그림 10>에 의하면 대체로 2009년을 기점

<그림 10> 남북갈등 관련 공기어의 변화



으로 남북갈등 관련 단어들이 크게 증가하고 있는 점이 주목된다.

한편 남북이 상대 집단을 비난할 때 쓰고는 했던 ‘괴뢰’라는 단어도 (12)~(13)에서 확인할 수 있듯이 2000년과 2010년도에는 그 쓰임에 큰 차이를 보이고 있다. 즉, 2000년에는 ‘괴뢰’라는 규정에 대한 용어 수정이 필요하다는 다소 긍정적인 문맥에서 ‘괴뢰’가 쓰이고 있는 반면 2010년에 이르러서는 상대 집단을 비방하기 위한 용법으로 쓰이고 있다.

(12) 이밖에 북한을 ‘괴뢰’로 규정한 ‘국가유공자예우에 관한 법률’ 등 10여 개 법률도 용어를 고쳐야 할 것으로 보인다(『한겨레신문』, 2000년 6월 21일).

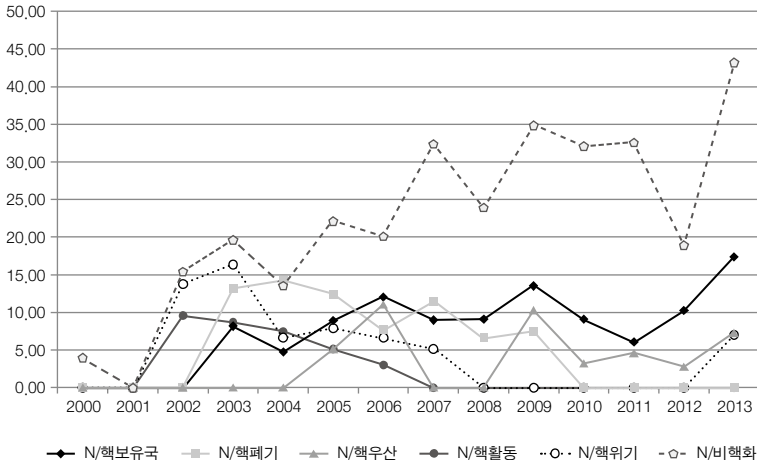
(13) 그러면서도 북한 방송은 세종시 문제 등을 거론하며 이명박 정부를 ‘괴뢰 당국’으로 비방했다(『중앙일보』, 2010년 1월 19일).

6) 핵 문제

북한의 핵 문제는 2000년 이후 가장 높은 관련성을 보인 공기어라는 사실은 <그림 3>에서도 지적된 바 있다. ‘핵’이 포함된 주요 공기어에는 ‘핵보유국, 핵폐기, 핵우산, 핵활동, 핵위기, 비핵화’ 등이 포함되는데 이들의 변화 추이는 <그림 11>과 같다.

<그림 11>에서 주목되는 추이는 연도별로 증감의 편차는 있으나 ‘비핵화’와 ‘핵보유국’의 변화 양상이다. 이들은 2002년과 2003년, 관련성이 급증한 이후 지속적으로 증가 추세를 보이고 있다. 한편 (14)~(15)에서도 확인되듯이 2002년의 ‘비핵화’와 2013년의 ‘비핵화’는 사용된 문맥에서 큰 차이를 보이고 있다. 즉, 2002년만 하더라도

<그림 11> 북핵 관련 공기어의 변화



‘비핵화’는 한반도의 비핵화를 추진하는 긍정적인 문맥에서 논의된 것이라면 2013년의 비핵화는 북한의 ‘비핵화 포기’와 같이 부정적인 상황에서 사용된 것이다. 한편 (16)에서 보이듯이 2003년에 이르러 이미 북한의 핵보유국 선언이 임박했음을 암시하는 기사도 나타나고 있으며 이는 실제로 2005년 북한의 핵보유국 선언으로 이어지게 된다.

(14) 중국은 한반도 비핵화 원칙에 따라 북한에 핵개발 포기를 요구 하겠다는 입장을 밝혔고, 일본도 ‘북한의 핵 포기 없이 수교는 없다’고 말했다(『조선일보』, 2002년 10월 21일).

(15) 특히 이날 북한은 유엔 안전보장이사회의 추가 제재에 반발해 비핵화 포기를 밝히면서 3차 핵실험 가능성으로 위협했다(『조선일보』, 2013년 1월 24일).

(16) 미 행정부의 일부 관리들은 “이미 북한이 핵보유국 선언 직전

단계까지 갔다”며 “북한이 핵보유국 선언을 해버리면 외교적 해결 가능성은 희박해질 것”으로 우려하고 있다(『동아일보』, 2003년 7월 23일).

7) 탈북자 문제

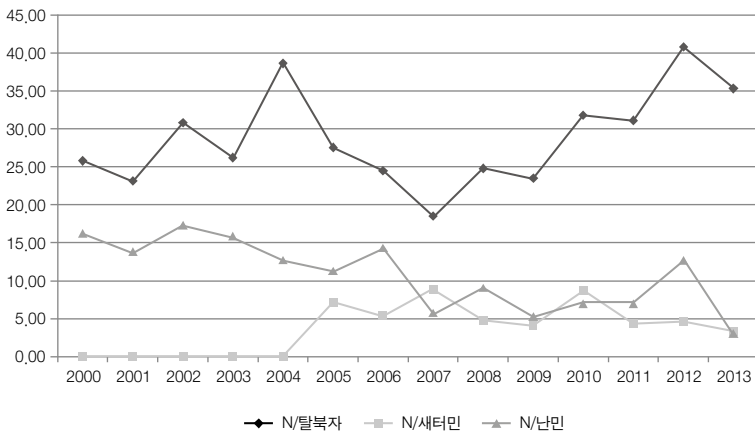
북한 사회에서 이탈하는 주민을 지칭하는 ‘탈북자’는 <그림 12>에서처럼 2004년과 2012년에 관련성이 가장 높은 것으로 나타났다.

특히 2012년에는 김정은 정권의 출범과 함께 북한과 중국이 탈북자 문제에 적극 개입함으로써 북한과 ‘탈북자’의 관련성이 크게 증가한 것으로 보인다.

(17) 탈북자들과 힘을 합쳐 ‘자유북한방송’이라는 이름으로 첫 인터넷 방송국을 마련하게 된 것이다(『조선일보』, 2004년 5월 14일).

(18) 북한도 김정은 체제 출범 이래 탈북자 차단에 더욱 안간힘을

<그림 12> ‘탈북자’ 관련 공기어의 변화



쓰고 있다(『동아일보』, 2012년 2월 20일).

이와 함께 2005년부터는 남한 통일부가 공식으로 규정한 ‘새터민’이라는 새로운 단어의 관련성이 등장하고 있다(19). 또한 북한 탈북자의 난민 지위 획득과 관련한 문제도 주요한 이슈로 꾸준히 등장하고 있다(20).

(19) 통일부는 지난 10여 년 동안 탈북자를 일컬었던 ‘북한이탈주민’이라는 용어를 ‘새터민’으로 점차 바꿔 나가기로 하고 이의 사용을 권장하기로 했다(『한겨레신문』, 2005년 1월 10일).

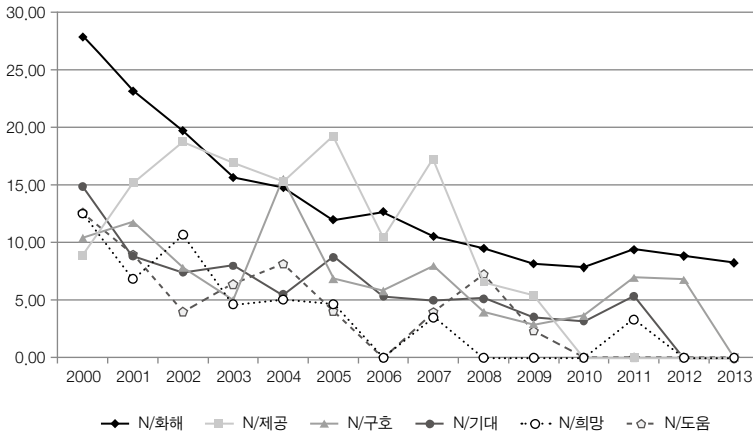
(20) 중국 내 북한 탈북자들을 정치적 난민으로 인정하기 위한 결의안 채택이 한국과 미국 의회에서 동시에 추진된다(『동아일보』, 2001년 7월 12일).

8) 남북문제의 변화, 긍정성의 감소

한편 북한과 관련한 공기어 가운데에는 뚜렷한 감소 추이를 보이는 공기어 집단을 확인할 수 있는데(<그림 13>), 이러한 단어들은 대체로 긍정적인 의미를 담고 있는 것으로 볼 수 있다.

즉, ‘화해, 제공, 구호, 기대, 희망, 도움’ 등과 같은 공기어는 구체적인 문맥에서는 일부 차이가 있을 수 있으나 이들 단어들이 전반적으로 긍정적인 의미로 쓰인다는 점은 부정하기 어렵다. 또한 이들 단어가 2000년대 후반 들어 관련성이 지속적으로 감소하고 있다는 점도 흥미로운 양상이다. 이들은 2000년대 들어 대북 관계가 점차 악화되어 온 상황을 반영하고 있는 것으로 해석된다. 이와 같이 공기어의 긍정성과 부정성에 대한 논의는 4절의 의미운율 분석에서 좀 더 구체적으로 검토될 것이다.

<그림 13> 남북문제 관련 공기어의 변화



4. ‘북한’의 의미운율 분석

의미운율(semantic prosody)이란 어떤 단어가 주변 단어와 함께 쓰이면서 발생하게 되는 의미적 아우라(aura)로 정의할 수 있다(Louw, 1993). 즉, 대상어가 특정한 부류의 의미적 속성을 공유하는 단어들과 반복적으로 쓰이면서 획득하게 되는 주변적 의미라 할 수 있는데 이때 의미적 속성은 대체로 긍정성과 부정성으로 구분된다.

이 연구에서는 2000~2013년 4개 일간지에서 추출한 ‘북한’의 공기어 가운데 관련성이 높은 공기어들(t 점수 100 이상으로 제한)을 의미운율에 따라 분류해 보았다. 즉, ‘북한’과 높은 관련성을 보인 515개의 명사들을 긍정성을 표현하는 것과 부정성을 표현하는 것으로 분류하였는데 그 결과는 다음과 같다.¹²⁾

[대응: 부정(98개)]

위협 도발 제재 중단 전쟁 공격 납치 붕괴 테러 압박 사태 우려 동결
살상 비난 긴장 침범 경고 무력 고립 사격 거부 통보 배치 비판 태세
지적 억류 일방적 포용 침투 방어 감시 망명 적대 타격 통제 이탈 독재
합정 차단 압력 남북 배제 분단 유감 처형 감행 선제공격 악화 봉쇄
강행 악 금지 규탄 기습 강제 나포 증강 공작 겨냥 폭파 비랑 추방 협박
중지 침몰 탈퇴 위기 저지 적대시 위반 강성 도발적 응징 대치 폭침
유린 강경책 위협적 경색 실패 공세 폭격 침략 해체 대결 침공 굶주림
굴복 불안정 충돌 참상 억제 탄압 호전적 책동 반발

[대응: 긍정(44개)]

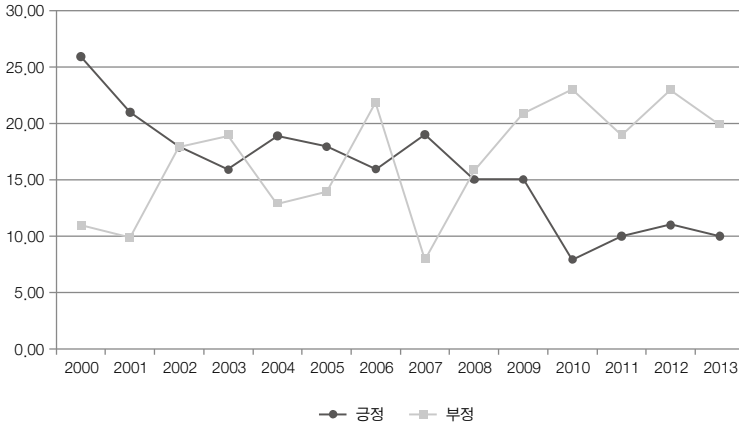
지원 합의 평화 해결 조치 협력 개방 개선 진전 보장 교류 이행 송환
인도적 안전 정상화 조처 평화적 해제 화해 동맹 인정 노력 개혁 공조
상봉 초청 복귀 원조 추진 합의서 대처 관심 협조 신뢰 불가침 타결
합의문 재건 번영 우호적 제공 성사 수용

이러한 긍정적 의미를 가진 공기어와 부정적 의미를 가진 공기어
들의 연도별 관련성 변화를 제시하면 <그림 14>와 같다.

한편 긍정성을 가진 공기어가 축소하는 데 비해 부정성을 가진 공
기어가 증가하는 구체적인 양상은 <표 3>에서도 구체적으로 확인
할 수 있다. <표 3>은 2000년대 초반, 중반, 후반에 해당하는 3개

12) 강범모(2011)에서도 논의된 바와 같이 명사의 의미운용은 중립적인 경우가 많고 문맥에 따라 긍정성이나 부정성이 달리 해석될 수 있다는 점에서 좀 더 신중한 해석을 필요로 한다. 그럼에도 불구하고 전반적인 경향성을 파악하는 데에는 충분히 의의가 있다는 점을 여기서는 중시하기로 한다.

<그림 14> 긍정/부정 공기어의 연도별 변화(세로축은 공기어의 종류 수)



<표 3> 2000년 vs. 2006년 vs. 2013년

2000년	긍정 (24개)	지원(14), 협력(23), 평화(27), 개선(35), 합의(36), 화해(59), 해결(95), 초청(99), 해제(123), 정상화(124), 긍정적(141), 완화(142), 포용(191), 성사(195), 재개(213), 환영(214), 이해(216), 햇볕(217), 원조(227), 인도적(231), 관심(239), 기대(240), 체결(249), 수용(282)
	부정 (11개)	테러(44), 위협(58), 납치(120), 전쟁(126), 제재(144), 중단(161), 긴장(189), 살상(267), 도발(274), 붕괴(293), 비난(296)
2006년	긍정 (14개)	복귀(30), 조치(31), 지원(48), 해결(50), 재개(93), 합의(103), 협력(116), 해제(134), 포용(160), 햇볕(184), 진전(239), 인도적(262), 평화적(266), 개선(296)
	부정 (31개)	제재(9), 위협(24), 압박(46), 강행(59), 동결(65), 중단(67), 사태(73), 살상(84), 우려(89), 붕괴(95), 전쟁(100), 납치(113), 도발(121), 경고(130), 고립(133), 불법(149), 강경(155), 위기(156), 무장(183), 거부(187), 공격(195), 비난(199)
2013년	긍정 (10개)	평화(37), 조치(53), 재개(76), 신뢰(85), 합의(107), 협력(114), 해결(158), 정상화(187), 진전(252), 지원(290)
	부정 (34개)	도발(4), 위협(9), 제재(17), 공격(49), 전쟁(54), 중단(65), 처형(71), 긴장(87), 압박(90), 타격(103), 사태(131), 비난(137), 강행(139), 숙청(140), 붕괴(144), 해킹(150), 폐쇄(157), 폭침(162), 경고(166), 종북(168), 일방적(176), 억류(179), 테러(207), 거부(214), 협박(215), 우려(222), 차단(228), 고립(235), 무력(236), 침투(265), 응징(268), 납치(273), 이탈(292), 선동(294)

연도의 긍정, 부정 공기어의 목록을 정리한 것이다(괄호 안의 숫자는 관련성 순위를 표시).

‘지원’과 ‘도발’이라는 공기어에 의해 단적으로 드러나듯이 2000년에는 긍정적 공기어가 높은 비중을 차지한 반면, 2013년에 와서는 부정적 공기어의 비중이 높아진다는 특성을 보인다. 즉, 2000년에는 긍정적인 공기어가 차지하는 비중이 부정적인 공기어의 약 두 배 이상 많을 뿐 아니라 관련성 순위도 긍정적 공기어가 부정적 공기어를 압도한다. 상위 100위권 안에 긍정적인 공기어는 8개인데 비해 부정적인 공기어는 2개에 불과하다. 특히 2013년의 ‘도발, 위협, 제재’ 등의 부정적 공기어의 관련성 순위가 매우 높다는 점이 주목된다.

5. 결론과 향후 과제

지금까지 신문 빅 데이터인 ‘물결21’ 코퍼스를 기반으로 하여 ‘북한’의 공기어들을 통계적으로 분석해 보았다. 본문에서 논의된 주요 내용을 간략히 요약하고 향후 과제를 제시하는 것으로 결론을 삼는다.

먼저 북한과 관련성이 높은 공기어들을 주제별로 분류하고 해당 주제에 포함된 공기어들의 관련성이 어떻게 변화하였는지를 관찰하였다. 이 주제에는 북한을 둘러싼 국제 정세의 변화, 남남갈등의 변화 추이, 북한의 생활상, 남북 지원 사업의 소멸, 남북갈등의 심화, 북핵 문제, 탈북 문제 등이 포함되었으며, 각 주제에 속하는 구체적인 공기어들의 관련성 변화를 통해 북한 관련 이슈의 변화 추이를 명시적으로 제시해 보았다. 또한 북한 관련 공기어의 의미운율을 긍정성과 부정성을 나타내는 공기어로 분류하고 이들의 관련성 변화를 통해

2000년대 초반에 비해 2000년대 중반 이후에는 부정성이 강조되는 변화를 보이고 있음을 제시하였다.

이와 같은 일련의 결과들은 어쩌면 북한 문제에 대한 기존의 정성적 분석과 크게 그 차이가 두드러지지 않게 보일 수도 있다. 그러나 본고에서 논의된 결과들은 주관이나 선입견이 배제된 빅 데이터에 기반을 둔 실제 단어 사용 양상을 계량적으로 분석한 결과라는 점, 구체적인 단어를 통해 북한 관련 이슈를 명시적으로 확인할 수 있다는 점 등에서 신문 빅 데이터에 기반을 둔 계량적, 통계적 접근이 가지는 의의는 적지 않을 것이다.

그럼에도 불구하고 본고는 여전히 남은 과제를 안고 있다. 시기를 확장하여 좀 더 거시적인 차원의 북한 관련 이슈의 변화를 살펴보거나 변화 추이를 분석할 수 있는 다양한 실험적 방법을 적용해 보는 등의 과제는 향후 연구로 남기기로 한다.

■ 접수: 3월 23일 / 수정: 4월 13일 / 채택: 8월 1일

참고문헌

1. 국내 자료

1) 단행본

김일환·정유진·강범모·김홍규, 『‘물결21’ 코퍼스의 구축과 활용』(서울: 소명출판, 2013).

에리즈 에이든·장바티스트 미셸, 『빅데이터 인문학: 진격의 서막』, 김재중 옮김 (과주: 사계절, 2015).

2) 논문

강범모, “형용사와 명사의 의미운율,” 『언어』, 제36권 1호(2011), 1~23쪽.

강정호, “빅 데이터를 이용한 선제적 사이버전 강화 방안 연구,” 『보안공학연구논문지』, 제13권 3호(2016), 195~204쪽.

김일환·이도길·강범모, “공기 관계 네트워크를 이용한 감정명사의 사용 양상 분석,” 『한국어학』, 제49권(2010), 119~148쪽.

남길임, “어휘의 공기 경향성과 의미적 운율,” 『한글』, 제298권(2012), 135~164쪽.

송정호·이상원, “빅데이터 기법을 활용한 북한의 권력구조 분석: 김정일, 김정은 시대 비교,” 북한연구학회 2015 세계북한학 학술대회 발표.

송태민, “소셜 빅데이터를 활용한 통일인식 동향 분석 및 예측,” 『북한경제리뷰』, 제17권 9호(2015), 1~8쪽.

이도길, “공기어로 살펴본 신문별 유사도 비교 연구: 조선일보, 동아일보, 중앙일보, 한겨레신문을 중심으로,” 『민족문화연구』, 제59권(2013), 613~635쪽.

이도길·최재웅, “물결 21 코퍼스: 공유와 확산: 물결21 코퍼스: 공개 웹 자원 및 활용 도구,” 『민족문화연구』, 제64권(2014), 3~23쪽.

최유숙, “신문기사에 나타난 북한이탈주민 지칭어 분석: 지칭어와 관련어의 공기어를 중심으로,” 『어문론집』, 제67권(2016), 33~66쪽.

2. 국외 자료

1) 논문

Baker, Paul, Costas Gabrielatos and Tony Mcenery, “Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word ‘Muslim’ in the British Press 1998-2009,” *Applied Linguistics*(2012), pp.1~25.

Louw, B., “Irony in the text or insincerity in the writer? the diagnostic potential of semantic prosodies,” in M. Baker, G. Francis and E. Tognini-Bonelli(eds.) *Text and Technology: In Honour of John Sinclair*(Amsterdam: John Benjamins, 1993), pp.157~176.

Sketching North Korea: The Representations of North Korea Based on Newspaper Big Data, ‘Trends 21’

Kim, Ilhwan(Korea University)·Lee, Do-gil(Korea University),
Kang, Jin Woong(Korea University)

This study aims to more objectively explore the images of North Korea represented in South Korean newspapers by quantitatively reviewing the patterns of language use related to North Korea in the recent 14 year (2000~2013) newspaper Big Data. In spite of a high profile of North Korea, existing studies had difficulty in analyzing the images of North Korea represented in South Korean mass media statistically and quantitatively. This article reveals the changing images of North Korea through analyzing the co-occurrence words and semantic prosody in the Trends 21 Corpus Big Data. In particular, it examines the changes in the international circumstances, domestic conflicts, the living conditions of North Korea, the South’s aid project for the North, the conflicts of the two Koreas, North Korean nuclear

issue, and the North Korean refugee problem through the co-occurrence word analysis. It also offers an analysis of South Korean mentality toward North Korea in terms of the semantic prosody.

Keywords: North Korea, newspaper Big Data, ‘Trends 21’ corpus, quantitative approach, statistical analysis, co-occurrence word, semantic prosody, patterns of language use