Analyses & Alternatives 2023 7(1): 115-145
DOI: 10.22931/aanda.2023.7.1.004

An Analysis of the 20th National Congress Report through Text-mining Methods*

Kwon, Dokyung · Kim, Jungsoo · Park, Jihyun**
Pusan National University, Jeonbuk National University, Green Carbon Research Center

Abstract

The 20th National Congress of the Chinese Communist Party (hereafter referred to as "the 20th National Congress") was under the global spotlight long before it was held for seven days from 16 to 22 October 2022. People wondered whether Xi Jinping would secure a third term as China's leader or whether he would lay the foundations to be in power forever during the third term. In Korea, the press and media questioned whether the event would become the "crowning of Emperor Xi (Xi Huangdi)," whose power rivaled that of the first emperor in China, Shi Hunagdi, and featured the scene where Hu Jintao was forced to leave the venue during the Congress. On the other hand, many Korean academics focused more on how Xi would organize the Politburo and its Standing Committee and whether the outline of his heirs would appear during the event. This tendency in academia in turn worsened the media's concerns.

This paper presents a quantitative analysis of the 20th National Congress Report, as opposed to an analysis of Xi's political intentions at the event. The National Congress Report outlines the Party's visions, goals, and strategies for the next five years in politics, economy, society, culture, foreign affairs, and relationship with Taiwan. The authoritative document is rich in narrative and logic and deserves academic study. This research analyzes the 18th, 19th, and 20th Reports by identifying their keywords and regular expressions and checking their frequency and percentage through text-mining methods. This approach enables the quantification and visualization of the significant changes in the Party's sovereign vision over the fifteen years of Xi's rule from 2013 to

* This research was supported by "Research Base Construction Fund Support Program" funded by Jeonbuk National University in 2021.

** Kim, Jungsoo. Jeonbuk National University. E-mail: jungkim@jbnu.ac.kr

www.kci.go.kr

2027.

Keywords
20th Party Congress reports, 18th Congress reports, 19th Party Congress reports, text mining, XiJinping

www.kci.go.kr

Analyses & Alternatives 2023 7(1): 115-145
DOI: 10.22931/aanda.2023.7.1.004

텍스트 마이닝을 활용한 중국공산당 20차 당대회 보고문 분석*

권도경 · 김정수*** · 박지현****
부산대학교 중어중문학과, 전북대학교 중어중문학과, 녹색탄소연구소

요 약

2022년 10월 16일부터 22일까지, 총 이레 동안 진행된 중국공산당 제20차 전국대표대회(이하 '20차 당대회'로 약칭)는 개최 훨씬 이전부터 '시진핑이 3연임을 할 것인가', '3연임을 함으로써 영구 집권의 기초를 닦을 것인가' 등의 문제를 둘러싸고 세간의 주목을 받았다. 한국 언론의 관심은 주로 20차 당대회가 진시황에 버금가는 '시황제의 대관식'이었다는 점에, 혹은 당대회 진행 중 전 총서기 후진타오(胡锦涛)가 강제퇴장 당하는 장면에 모아졌고, 학계도 20차 당대회에서 시진핑이 정치국 위원과 상무위원에 대한 인사 배치를 어떻게 할 것인지에, 또 시진핑의 후계 구도가 드러날 것인가 등에 집중적인 관심을 기울이면서 오히려 언론의 우려를 강화시켰다.

본 논문은 20차 당대회에서의 시진핑의 정치적 의도에 대한 해석과는 거리를 두면서, 당대회 보고문에 대한 계량적 분석을 시도할 것이다. 중국공산당의 당대회 보고문은 향후 5년 간의 정치, 경제, 사회, 문화, 외교 및 양안관계, 과학기술 등의 분야에 대한 중국공산당의 통치 비전 및 목표, 그리고 전략적 방향을 담고 있으면서 그 자체로 높은 서사성과 논리성을 가지고 있는 문건으로 학술적 가치가 높기 때문이다. 본 논문은 당대회 문건을 대상으로 텍스트 마이닝 방법론을 사용하여 주요 어휘 빈도수 조사 및 분석, 키워드 분석, 주요 표현 조사 및 분석을 진행할 것이다. 이를 통해 18차 당대회에서 20차 당대회까지, 즉 2013년부터 2027년까지 15년의 시진핑 집권기 동안 중국공산당 통치 비전의 거시적인 변화를 계량화, 시각화할 것이다.

* 이 논문은 2021년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.
** 교신저자
*** 제1저자: 전북대 중어중문학과/전북대 중국·아시아연구소 소장. 이메일: jungkim@jbnu.ac.kr
**** 공동저자

www.kci.go.kr

주제어
20차 당대회 보고문, 19차 당대회 보고문, 18차 당대회 보고문, 텍스트 마이닝, 시진핑

www.kci.go.kr

Ⅰ. 서론

2022년 10월 16일부터 22일까지, 총 이레 동안 진행된 중국공산당 제20차 전국대표대회(이하 '20차 당대회'로 약칭)는 개최 훨씬 이전부터 '시진핑(習近平)이 3연임을 할 것인가', '3연임을 함으로써 영구 집권의 기초를 닦을 것인가' 등의 문제를 둘러싸고 세간의 주목을 받았다. 한국 언론의 관심은 주로 20차 당대회가 진시황에 버금가는 '시황제의 대관식'이었다는 점에, 혹은 당대회 진행 중 전 총서기 후진타오(胡锦涛)가 강제퇴장 당하는 장면에 모아지면서, 실제 20차 당대회가 중국 사회의 현재와 미래에, 그리고 더 나아가 한국을 비롯한 전 세계의 앞날에 가져올 영향과 변화에 대해서는 주의를 덜 기울였다. 이런 사정은 학계도 마찬가지다. 20차 당대회에서 시진핑이 정치국 위원과 상무위원에 대한 인사 배치를 어떻게 할 것인지에, 또 시진핑의 후계구도가 드러날 것인가 등에 집중적인 관심을 기울이면서 오히려 언론의 우려를 강화시켰다.

본 논문은 20차 당대회에서의 시진핑의 정치적 의도에 대한 해석과는 거리를 두면서, 당대회 보고문에 대한 계량적 분석을 시도할 것이다. 중국공산당의 당대회 보고문은 향후 5년 간의 정치, 경제, 사회, 문화, 외교 및 양안관계, 과학기술 등의 분야에 대한 중국공산당의 통치 비전 및 목표, 그리고 전략적 방향을 담고 있으면서 그 자체로 높은 서사성과 논리성을 가지고 있는 문건으로 학술적 가치가 높다. 문건 자체에 대한 계량적 분석이 때로는 객관적 해석을 도출하는 데 도움이 될 수 있을 것이라 생각한다. 본 논문은 당대회 문건을 대상으로 텍스트 마이닝 방법론을 사용하여 계량적 분석을 시도할 것이다. 이를 통해 18차 당대회에서 20차 당대회까지, 즉 2013년부터 2027년까지 15년 시진핑 집권기 동안 중국공산당 통치 비전의 거시적인 변화를 계량화, 시각

www.kci.go.kr

화할 수 있을 것이라 기대한다.

II. 연구 설계

1. 선행 연구

2022년 가을에 치러진 20차 당대회의 주요 쟁점은 인사와 노선으로 정리된다(양갑용, 2022). 당대회 개최 이전부터 시진핑 지도체제와 인사, 대외관계의 변화 가능성에 이목이 집중되었던 만큼, 당대회가 끝난 이후에도 주로 이런 문제들을 중심으로 연구가 이어지고 있다. 양갑용과 주장환은 주로 20차 당대회 이후 시진핑 집중지도체제와 그 인사의 결과에 대해 정치학의 관점에서 정리, 전망하고 있다. 그리고 표나리와 박소희는 외교의 관점에서, 특히 미중 대결 국면에서 20차 당대회가 현 미중관계와 국제관계에 미칠 영향에 대해 해석, 전망하고 있다.

국내에서 정치와 외교의 관점에서의 분석을 제외하고는 20차 당대회에 대한 분석은 거의 없는 편이다. 특히 20차 당대회 보고문 자체에 대한 계량적 분석은 찾아볼 수 없다. 본 논문은 20차 당대회 보고문인 중국특색 사회주의의 위대한 기치를 높이 들고 사회주의 현대화 국가의 전면 건설을 위해 단결분투하자(高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗)를 주요 분석 대상으로 삼아 텍스트 마이닝 방법론을 통해 어휘 빈도 분석, 키워드 분석, 주요표현 분석 등 계량적인 분석을 시도할 것이다. 동시에 이를 시진핑의 집권기에 속한 18차, 19차 당대회 보고문과 비교함으로써 20차 당대회 보고문의 특징을 부각시킬 것이다.

www.kci.go.kr

2. 분석 대상

본 논문은 20차 당대회 보고문 자체를 분석의 대상으로 삼는다. 또 20차 당대회 보고문의 내용을 특징화하기 위해 이를 18차, 19차 당대회 보고문과 동일한 조건에서 비교 분석할 것이다. 따라서 본 논문이 분석 대상으로 삼는 문건은 18차 당대회 보고문 '중국특색사회주의의 길을 따라 흔들림 없이 전진하고 소강사회의 전면적 완성을 위해 분투하자(坚定不移沿着中国特色社会主义道路前进 为全面建成小康社会而奋斗)', 19차 당대회 보고문 '소강사회의 전면적 완성에서 승리하고 신시대 중국특색 사회주의의 위대한 승리를 거머쥐자(决胜全面建成小康社会 夺取新时代中国特色社会主义伟大胜利)', 20차 당대회 보고문 '중국특색 사회주의의 위대한 기치를 높이 걸고 사회주의 현대화 국가의 전면적 건설을 위해 단결해서 분투하자'이다.

3. 분석 절차

1단계) 원 텍스트 추출

본 연구는 당대회 보고문 분석을 위해 중국공산당원(中国共产党员网) 공식 홈페이지(https://www.12371.cn/)에 게재된 문건을 사용했다. 개요는 아래 [표 1]과 같다.

[표 1] 18차~20차 당대회 보고문 개요

20차 당대회 보고문	
제 목	高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗—在中国共产党第二十次全国代表大会上的报告
발표자	习近平
발표일	2022년 10월 16일

www.kci.go.kr

링 크	https://www.12371.cn/2022/10/25/ARTI1666705047474465.shtml
19차 당대회 보고문	
제 목	决胜全面建成小康社会 夺取新时代中国特色社会主义伟大胜利—在中国共产党第十九次全国代表大会上的报告
발표자	习近平
발표일	2017년 10월 18일
링 크	https://www.12371.cn/2017/10/27/ARTI1509103656574313.shtml
18차 당대회 보고문	
제 목	坚定不移沿着中国特色社会主义道路前进 为全面建成小康社会而奋斗—在中国共产党第十八次全国代表大会上的报告
발표자	胡锦涛
발표일	2012년 11월 8일
링 크	https://www.12371.cn/2012/11/17/ARTI1353154601465336.shtml

중국공산당원 공식 홈페이지에 게재되어 있는 이 세 개의 문건을 사용해 세 개의 원 텍스트 파일을 만들었다.

2단계) 품사 분석 및 데이터 정제

다음으로 이렇게 만들어진 세 개의 파일 속 모든 어휘에 대해 품사 분석을 실행했다. 품사 분석을 위해서 베이징사범대학교(北京师范大学)에서 제공하는 CorpusWordParser라는 품사 분석 프로그램을 사용했으며, 각 단어 뒤에 품사 태그가 표시된 품사 분석 파일을 얻었다. 그 후 CorpusWordParser의 품사 분석 결과를 좀 더 정교하게 하기 위해 분석 결과의 내용을 일일이 살펴보며 교정하였다. CorpusWordParser의 품사 분석 결과에서 자주 보이는 오류로는 끊어 읽기 오류, 신조어 인식 오류, 다의어 품사 오류 등이 있었다. 필자가 수정한 어휘는 Emeditor 프로그램을 이용해 별도로 목록으로 만들어, 세 개 파일의 품사 분석에

www.kci.go.kr

대한 수정, 변경에 동일하게 적용하였다. 그리하여 최종적으로 아래 [표 2]와 같은 당대회 보고문 텍스트를 만들었다.

[표 2] 당대회 보고문 총 어휘수

18차 당대회		19차 당대회		20차 당대회	
총 어휘 종류 (type)	2527	총 어휘 종류 (type)	3012	총 어휘 종류 (type)	1775
총 어휘 누적 (token)	13725	총 어휘 누적 (token)	15232	총 어휘 누적 (token)	6854

CorpusWordParser가 구분하는 35개의 품사 중 본 논문의 당대회 보고문 분석에 의미 있다고 판단되는 실사 중심의 어휘를 추출해 분석용 어휘 목록을 만들었다. 최종적으로 분석 파일에는 명사, 동사, 형용사, 처소명사, 지명, 기구 명칭, 관용어 및 축약어 등 여덟 종류의 어휘가 포함되었다.

[표 3] 분석 대상 품사 선택

n	名词	nt	时间名词	nd	方位名词	nl	处所名词
nh	人名	nhf	姓	nhg	名	ns	地名
ni	机构名	nz	其他专名	v	动词	vd	趋向动词
vl	联系动词	vu	能愿动词	a	形容词	f	区别词
m	数词	q	量词	mq	数量结构	d	副词
r	代词	p	介词	c	连词	u	助词
e	叹词	o	拟声词	i	习用语	j	缩略语
h	前接成分	k	后接成分	g	语素字	x	非语素字
w	标点符号	ws	非汉字字符	wu	其他符号		

www.kci.go.kr

3단계) 어휘 데이터 목록 확보

원 텍스트에 대해 품사 분석 및 데이터 정제, 그리고 불용자 제거 및 실사 중심의 분석 대상 어휘 추출을 진행해 본 연구에서 분석 대상으로 삼을 어휘 데이터를 얻었다. 18차 당대회 보고문에서는 총 2,299개의 어휘, 총 10,546회의 누적 빈도수를, 19차 당대회 보고문에서는 총 2,754개의 어휘, 총 11,536회의 누적 빈도수를, 20차 당대회 보고문에서는 총 1,569개의 어휘, 총 5,077회의 누적 빈도수를 보였다. 결과는 아래 [표 4]와 같다.

[표 4] 당대회 보고문 분석 데이터 개요

18차 당대회			19차 당대회			20차 당대회		
태그	품사	빈도	태그	품사	빈도	태그	품사	빈도
n	名词	4458	n	名词	4831	n	名词	2056
v	动词	4011	v	动词	4329	v	动词	1803
a	形容词	1443	a	形容词	1481	a	形容词	746
i	习用语	431	i	习用语	601	i	习用语	346
j	缩略语	78	ns	地名	153	ns	地名	71
ns	地名	75	j	缩略语	93	j	缩略语	25
nl	处所名词	37	nl	处所名词	30	nl	处所名词	13
ni	机构名	13	ni	机构名	18	ni	机构名	15

총 어휘 종류	2,299	총 어휘 종류	2,754	총 어휘 종류	1,569
총 누적 (tokens)	10,546	총 누적 (tokens)	11,536	총 누적 (tokens)	5,077

4단계) 어휘 빈도, 키워드 및 주요 표현(N-gram) 추출

3단계에서 확보한 최종의 분석 데이터를 대상으로 해서 본 논문은 1) 어휘 빈도수 2) 키워드 3) 주요 표현(N-gram)을 추출, 분석함으로써

www.kci.go.kr

20차 당대회 보고문의 특징을 드러낼 것이다. 각 분석의 특징은 아래와 같다.

1) 어휘 빈도 분석: 특정 문서 내에서 자주 사용되는 어휘가 사용되는 빈도수에 따라 중요도를 분석하는 방법. 문서의 주제 및 내용을 계량화하는 데에 용이하다.

2) 키워드 분석: 참조 코퍼스 어휘와의 비교를 통해 특정 문건에서 특별하게 자주 쓰인(혹은 쓰이지 않은) 어휘를 추출(Laurence Anthony, 2012)해 특정 문건, 특정 어휘의 상대적인 중요도를 분석하는 방법. 어휘 빈도 분석으로 놓치지 쉬운 키워드를 추출, 분석할 수 있다.

본 논문에서 사용한 참조 코퍼스는 정치, 경제, 사회, 문화, 수필, 소설 등 각 분야의 다양한 문체를 포괄하고 있는 15 종류의 중국어 코퍼스다. 구체적인 내용은 아래 [표 5]와 같다.

[표 5] 키워드 추출을 위한 참조 코퍼스

코드	유형	분량(단어)	코드	유형	분량(단어)
A	신문 보도	165,259	J	학술 및 과학기술	279,169
B	신문 사설	113,368	K	보통 소설	92,584
C	잡지 평론	60,662	L	탐정 소설	106,428
D	종교	60,169	M	SF 소설	19,633
E	일상과 소비 취미	122,258	N	서스펜스 소설	111,744
F	통속적인 글	175,903	P	로맨스 소설	96,002
G	전기 및 회상록	296,505	R	코미디와 유머	31,081
H	정부 혹은 기구 공문	115,570			

3) 주요 표현 분석: 주요 표현 분석은 전산언어학의 N-gram을 응용한 분석이다. N-gram은 ‘주어진 텍스트 또는 음성 샘플에서 n개 항목

www.kci.go.kr

의 연속적인 시퀀스'를 가리키는 개념으로, 응용 프로그램에 따라 음소, 음절, 문자, 단어 등이 기본 쌍이 될 수 있다.(wikipedia에서 'N-gram' 검색). N-gram 분석을 이용하면 어휘의 조합으로 이루어진 주요 표현을 추출할 수 있다. 본 논문에서는 2개의 어휘로 이루어진 어휘 조합, N-gram(n=2)을 분석 대상으로 삼을 것이다.

각 당대회 보고문의 어휘 빈도 분석, 키워드 분석 및 주요 표현 분석을 위해서 AntConc 3.5.9 윈도우 버전을 사용했다.

Ⅲ. 연구 결과

1. 어휘 빈도 분석

각 당대회 보고문에 대해 제일 먼저 어휘 빈도수 조사를 진행했다. 먼저 각 당대회 보고문의 총 어휘 종류의 빈도수 상위 1%에 해당하는 어휘를 비교해 보자. 18차 당대회 보고문의 2,299개의 총 어휘중 상위 1%의 주요 어휘 22개, 19차 당대회 보고문의 2754개의 총 어휘중 상위 1%의 주요 어휘 27개, 20차 당대회 보고문의 1569개의 총 어휘중 상위 1%의 주요 어휘 15개의 주요 어휘를 표로 만들면 아래 [표 6]과 같다.

[표 6] 당대회 보고문의 주요 어휘 빈도수

차수	18차		19차		20차	
순위	어휘	빈도	어휘	빈도	어휘	빈도
	총 어휘 누적	10546	총 어휘 누적	11536	총 어휘 누적	5077
1	发展	266	发展	213	发展	97
2	建设	187	党	183	坚持	92
3	人民	124	人民	162	人民	78
4	党	108	建设	158	国家	70

www.kci.go.kr

5	坚持	108	坚持	132	党	69
6	社会	104	国家	106	推进	55
7	制度	87	实现	85	建设	54
8	经济	83	制度	83	全面	53
9	中国特色社会主义	81	社会	82	安全	49
10	文化	79	推进	81	体系	48
11	加强	77	体系	76	新	46
12	推进	76	加强	72	现代化	42
13	体系	73	政治	71	战略	38
14	提高	72	中国特色社会主义	70	中国	33
15	新	72	中国	69	加强	30
16	改革	67	全面	68		
17	社会主义	66	文化	63		
18	国家	64	改革	61		
19	基本	60	经济	58		
20	全面	55	完善	56		
21	创新	55	安全	54		
22	完善	55	领导	54		
23			创新	52		
24			我国	50		
25			伟大	48		
26			新	48		
27			社会主义	48		
…	…		…		…	

한정된 지면 내에서 사용 빈도가 높은 어휘는 물론 그 문건에서 매우 중요하게 다뤄지는 말일 것이다. 이상 상위 1% 어휘의 빈도수 결과에서 가장 눈에 띄는 점은 각 당대회 보고문에서 ‘发展’이라는 말이 가장 높은 빈도수를 보인다는 점이다. 미중 대결 등 국내외 이슈로 시진핑 독주 체제가 안팎으로 공고해지기 시작한 19차, 20차 당대회 보고문에서도 ‘发展’이 가장 높은 빈도수로 사용되고 있다. 이는 개혁개방

www.kci.go.kr

이후 중국의 발전주의 노선이 변함없다는 확실한 증거가 된다.

다만 빈도수 1위인 '发展'이라는 어휘의 총 어휘 누적에서의 점유율은 차이를 보였다. 18차 당대회 보고문에서 '发展'의 빈도는 266이고 총 어휘 누적수에서 2.5% 점유한다. 19차 당대회 보고문에서 '发展'의 빈도는 213이고 총 어휘 누적수에서 1.8% 점유한다. 20차 당대회 보고문에서 '发展'의 빈도는 97이고 총 어휘 누적수에서 1.9%의 점유한다. '发展'이라는 어휘는 18차, 19차, 20차 당대회 보고문에서 모두 가장 높은 빈도수를 보이지만, 문건에서의 비중은 18차에서 가장 높은 것으로 나타났다.

그렇다면 이 '发展'이라는 말이 각 당대회 보고문에서 모두 같은 맥락으로 쓰였을까? 이 문제를 해결하기 위해 '发展'이라는 어휘 앞에 꾸미는 말이 붙는 '○○+发展' 형식의 표현을 조사했다. 다양한 '○○+发展' 형식의 표현 중, 개념어를 형성할 수 있는 '○○(형용사어)+发展(명사형)'과 '○○(명사어)+发展(명사형)', '○○(형용사어)+发展(동사형)', '○○(동사어)+发展(동사형)' 등의 조합을 모두 추출했으며, 그 결과는 아래 [표 7]과 같다.

[표 7] 당대회 보고문에서 '○○+发展' 표현 출현 빈도

차수	18차		19차		20차	
구분	어휘	빈도	어휘	빈도	어휘	빈도
어휘	发展	266	发展	213	发展	97
형식	○○(n)+发展(n) ○○(a)+发展(n) ○○(a)+发展(v) ○○(v)+发展(v)	98	○○(n)+发展(n) ○○(a)+发展(n) ○○(a)+发展(v) ○○(v)+发展(v)	99	○○(n)+发展(n) ○○(a)+发展(n) ○○(a)+发展(v) ○○(v)+发展(v)	47

www.kci.go.kr

중국어에서 '发展'은 명사형으로도 동사형으로도 쓰일 수 있는데, 18차에서 20차 보고문에서 필자가 추출한 '○○+发展' 형식의 빈도는 대략 50% 내외로 집계됐다. '○○(명사형)+发展(명사형)'으로 쓰인 경우는 '经济+发展', '社会+发展', '科学+发展', '区域+发展' 등의 표현이 있다. '○○(형용사)+发展(명사형)'으로 쓰인 경우는 '高质量+发展', '创新+发展', '和平+发展' 등의 표현이 자주 쓰였다. '○○(형용사)+发展(동사형)'의 경우는 '健康+发展', '巩固+发展' 등의 표현이 자주 쓰였다. '○○(동사)+发展(동사형)'의 경우는 '加快+发展', '坚持+发展', '协调+发展' 등이 자주 쓰였다. 그 중 상위 10%의 표현을 [표 8]로 정리했다.

[표 8] 각 당대회 보고문 중 '○○+发展' 형식의 주요 표현

차수	18차		19차		20차	
구분	어휘	빈도	어휘	빈도	어휘	빈도
총 누적	发展	266	发展	213	发展	97
형식	○○+发展	98	○○+发展	99	○○+发展	47
1	经济+发展	16	坚持+发展	11	新+发展	10
2	和平+发展	15	和平+发展	10	高质量+发展	8
3	坚持+发展	11	巩固+发展	6	和平+发展	5
4	科学+发展	11	加快+发展	5	坚持+发展	5
5	社会+发展	10	中国+发展	5		
6	改革+发展	8	协调+发展	5		
7	健康+发展	6	我国+发展	5		
8	共同+发展	6	经济+发展	4		
9	区域+发展	6	事业+发展	4		
…	…		…		…	
			高质量+发展	1		

www.kci.go.kr

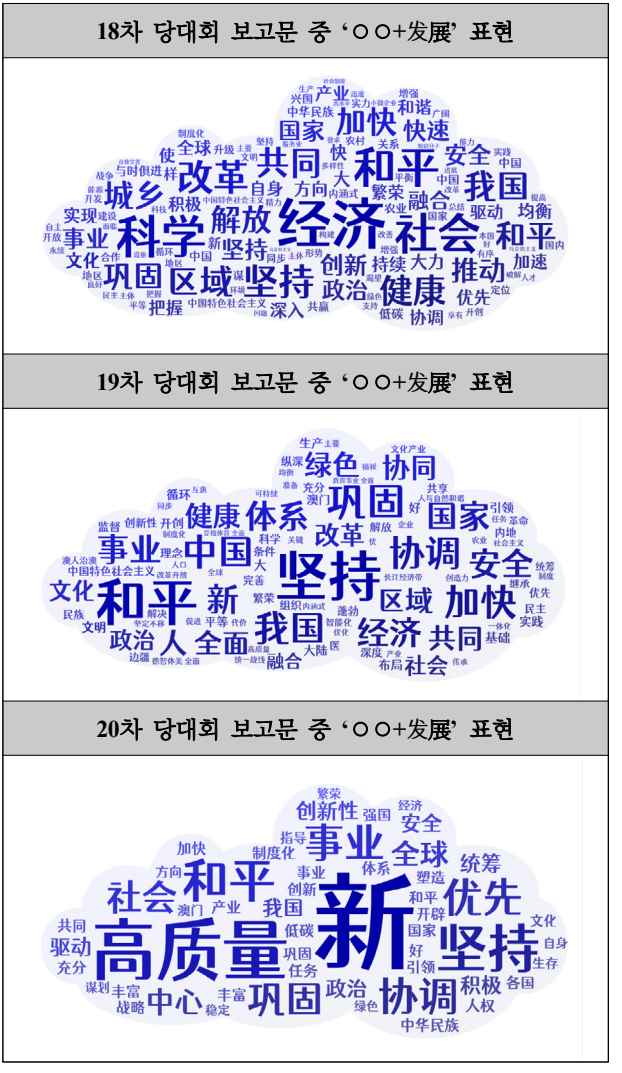
매우 흥미롭게도 각 당대회 보고문에서 다른 조합의 ‘○○+发展’이 사용된 것으로 나타났다. 18차 당대회 보고문에서는 ‘○○(명사형)+发展(명사형)’ 형식의 표현이 높은 빈도수를 보인다. ‘经济+发展’ 16회, ‘科学+发展’ 11회, ‘社会+发展’ 10회, ‘区域+发展’ 6회 등 다양한 분야에서의 발전을 가리키는 표현이 비교적 고른 빈도로 사용됐다. 그에 비해 19차 당대회 보고문에서는 ‘○○(형용사형)+发展(동사형)’과 ‘○○(동사형)+发展(동사형)’ 형식의 표현이 높은 빈도수를 보인다. ‘坚持+发展’ 11회, ‘和平+发展’ 10회, ‘巩固+发展’ 6회, ‘加快+发展’ 5회, ‘协调+发展’ 5회 등으로 주로 사용되었고, ‘○○(명사어)+发展(명사형)’ 형식으로는 ‘中国+发展’, ‘我国+发展’의 표현이 각각 5회의 빈도수로 ‘经济+发展’(4회)이라는 표현보다 많이 쓰였다. 20차 당대회 보고문에서는 새로운 개념어라 할 수 있는 ‘新+发展’과 ‘高质量+发展’이 각각 10회와 8회의 빈도수로 가장 빈번하여 사용됐다. 이 ‘新+发展’과 ‘高质量+发展’ 두 표현의 빈도수 합은 20차 당대회 보고문에 사용된 전체 ‘○○+发展’ 표현에서 38%를 점하며, 전체 N-gram(n=2)의 주요 표현 5077개 중 3%를 차지하는 수치로 상당히 중점적으로 사용되었음을 알 수 있다. 어휘의 조합 및 사용으로만 보면, 18차 당대회 보고문에서는 주로 발전의 내용이 중점이 되었다면, 19차 당대회 보고문에서는 ‘굳건하게’, ‘빠르게’, ‘공고하게’ 등이 자주 사용되면서 발전의 양태가 중점이 되었다. 20차 당대회에서는 ‘새로운 발전’, ‘높은 품질의 발전’ 등 아주 새로운 발전의 개념이 중점적으로 제시되고 있다. 20차 당대회 보고문에서 새롭게 제시된 이런 발전의 개념이 새로운 내용을 담보하지 않는다면 구호로 그치기 십상일 것이다. 이에 대해서는 향후 보다 심층적인 논의가 필요하다.

‘○○+发展’ 표현의 전체 분포를 보기 위해 이를 워드 클라우드로 시각화했다. 워드 클라우드 시각화를 위해 워드아트(www.wordart/create) 프로그램을 사용했고, 글자의 크기는 칼럼 값(주요 표현 빈도수)을 비

www.kci.go.kr

례하도록 설정했다.

[그림 1] 당대회 보고문 중 '○○+发展' 표현 시각화



www.kci.go.kr

[그림 1]을 통해 20차 당대회 보고문에서 ‘○○+发展’ 표현이 ‘新+发展’과 ‘高质量+发展’으로 집중되고 있으며, 또한 그 집중도가 18차와 19차 빈도수 1,2위 표현에 비해 매우 높음을 직관적으로 알 수 있다. 흥미로운 것은 빈도수 1,2 위의 ‘新+发展’과 ‘高质量+发展’은 모두 최근 새로 제기된 개념이라는 점이다. 빈도수 10의 ‘新+发展’은 20차 당대회 보고문에서 처음 사용된 말이며, 빈도수 8의 ‘高质量+发展’의 표현 역시 19차 당대회 보고문에서 처음으로 등장, 단 한 번 사용되었다. ‘高质量 发展’은 이후 2018년 1월, 중공중앙정치국위원이자 중앙재경영도소조판공실 주임인 류허(刘鹤)가 다보스 포럼에서 ‘고품질 발전을 추동해 전지구적 경제의 번영과 안정을 공동으로 촉진시키자’는 제목으로 연설을 해 중국 내외로 널리 알린 개념이기도 하다. 그는 이 연설에서 ‘중국이 이미 고속 성장의 단계를 지나 고품질 발전 단계로 전환되었다’고 천명하고 이것이 지금 중국이 당면한 총체적 요구라고 진단했다. 20차 당대회 보고문은 5년 전 제기된 이 ‘고품질 발전’의 총체적 요구를 다시 한번 확인하고 있다. 그렇다면 ‘新 发展’은 또 무엇일까? 그 내용을 적실하게 제시하지 못한다면 이 개념은 구호로 그치기 십상일 것이다.

2. 키워드 분석

한 문건에서 특정 어휘의 빈도수는 물론 중요성을 나타내는 지표 중 하나이다. 하지만 빈도수를 단순하게 비교하는 것으로는 한 문건에서 각별히 중요하게 쓰인 어휘 및 그 맥락을 특징화하기 어렵다. 여기에서는 각 당대회 보고문에서 특별히 중요하게 쓰인 어휘를 특징화하고자 한다. 이를 위해 AntConc 프로그램 내의 키워드 기능을 사용했다.

AntConc를 이용해 키워드를 추출한 결과, 18차 당대회 보고문의 경우 총 394종의 키워드 총 누적빈도수 6,459회, 19차 당대회 보고문의

www.kci.go.kr

경우 총 361종 키워드의 총 누적빈도수 6,540회 , 20차 당대회 보고문의 경우 총 265종 키워드에 총 누적빈도수 2,782회로 집계됐다.

[표 9] 당대회 보고문에서의 키워드 집계

차수	18차		19차		20차	
키워드	총 종류	394	총 종류	361	총 종류	265
	총 누적빈도	6,459	총 누적빈도	6,540	총 누적빈도	2,782

위의 결과에서 키워드 핵심도(keyness)를 기준으로 상위 5%에 해당하는 키워드를 정리했다. 핵심도란 '참조 말뭉치에서의 빈도와 비교할 때 목표 텍스트에서 어휘의 빈도'(Heather Froehlich, 2015)를 말한다. 편의상 아래 표의 핵심도 수치는 소수점 이하 생략하지만, 실제 분석에서는 소수점 이하 2자리까지의 수치를 사용했다.

[표 10] 당대회 보고문의 키워드(상위 5%)

차수	18차		19차		20차	
구분	키워드	핵심도	키워드	핵심도	키워드	핵심도
1	发展	754	党	892	坚持	539
2	建设	752	人民	687	人民	353
3	中国特色 社会主义	748	坚持	642	党	328
4	坚持	508	中国特色 社会主义	634	中国特色 社会主义	310
5	人民	490	建设	563	新时代	310
6	党	448	发展	489	全面	267
7	制度	316	新时代	326	推进	260
8	推进	297	推进	312	社会主义 现代化	256

www.kci.go.kr

9	体系	293	实现	307	国家	242
10	加强	284	体系	298	发展	231
11	社会主义	282	国家	286	体系	224
12	民主	249	制度	280	安全	205
13	完善	235	全面	266	战略	196
14	提高	232	加强	245	…	
15	增强	217	中华民族 伟大复兴	244		
16	健全	208	全党	244		
17	全面	204	法治	235		
18	加快	198	完善	231		
19	基本	195	…			
	…					

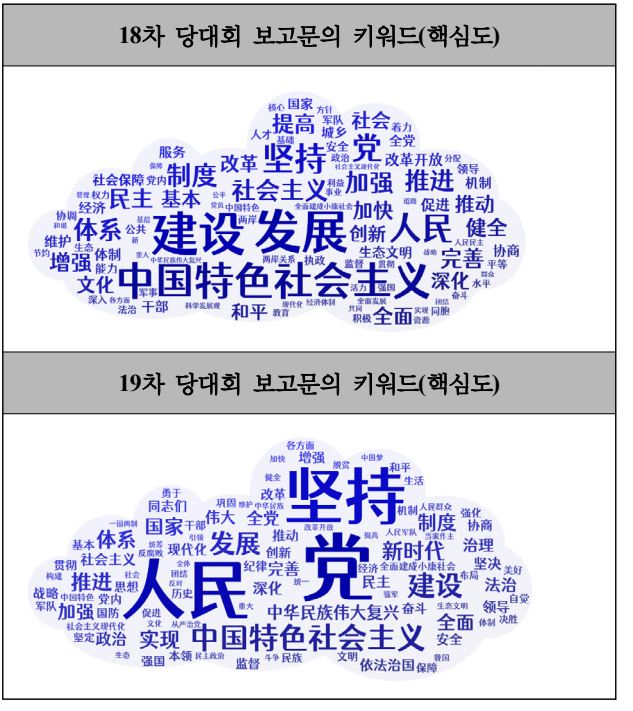
핵심도를 기준으로 한 키워드 조사에서 아주 흥미로운 점이 눈에 띈다. 앞서 어휘 빈도 조사에서 18차, 19차, 20차 당대회 보고문에서 '发展'이라는 어휘가 공히 1위를 기록했다는 점을 지적했다. 그런데, 핵심도 기준 키워드에서는 좀 다른 양상이 나타난다. 18차 당대회 보고문의 키워드는 여전히 빈도수 1위인 '发展'이 핵심 키워드 1위로 나타났지만, 19차 당대회 보고문에서는 '党'(핵심도 892)이, 20차 당대회 보고문에서는 '坚持'(핵심도 539)가 핵심 키워드 1위로 나타났다.

당대회 보고문에 대한 키워드 분석의 결과는 몇 가지 시사점을 준다. 먼저 AntConc의 키워드 분석은, 앞서 언급했듯이, 참조 코퍼스와 비교할 때 목표 텍스트에서 각별하게 높은 빈도로 사용된 어휘를 보여주는 기능이다. 이 말은 참조 코퍼스에서의 어휘 빈도수가 목표 텍스트의 키워드 분석에 직접적으로 영향을 준다는 말이기도 하다. 18차, 19차, 20차 당대회 보고문에서 '发展'이라는 말이 공히 빈도수 1위를 기록했음

www.kci.go.kr

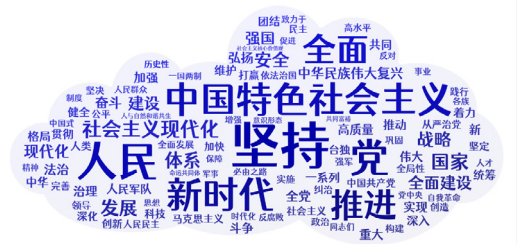
에도 19차, 20차 당대회 보고문의 핵심도 키워드가 '发展'이 아닌 것은 참조 코퍼스 자체에 '发展'이라는 어휘의 빈도가 높을 가능성을 시사한다. 둘째, 이를 적용해 다시 결과를 해석하면 19차 당대회 보고문에서는 다른 참조 코퍼스들에 비해 '党'이라는 어휘가, 20차 당대회에서는 '坚持'라는 어휘가 특별하게 빈번하게 쓰였음을 말해준다. 단순히 빈도수가 높기만 한걸까? 주요 키워드 핵심도의 변화 및 그 크기를 시각화하기 위해 각 당대회 보고문 키워드 100위까지의 어휘를 워드 클라우드로 시각해했다. 워드 클라우드 속 어휘는 중복되지 않게, 글자의 크기는 칼럼 값(키워드의 핵심도)을 따랐다.

[그림 2] 당대회 보고문의 키워드(핵심도) 시각화



www.kci.go.kr

20차 당대회 보고문의 키워드(핵심도)



[그림 2]의 워드 클라우드의 결과를 볼 때, 18차 당대회 보고문의 키워드는 ‘发展’이라는 어휘가 중심이지만, ‘发展’을 둘러싸고 ‘建设’, ‘中国特色社会主义’, ‘人民’, ‘党’, ‘制度’ 등의 어휘가 비슷한 크기(즉 고른 핵심도)를 보인다. 이와 비교할 때, 19차 당대회 보고문에서는 ‘党’을 중심으로 ‘人民’과 ‘坚持’라는 말이 두드러지게 큰 크기로 나타나고, 20차 당대회 보고문 역시 ‘坚持’를 중심으로 ‘人民’, ‘中国特色社会主义’, ‘党’, ‘新时代’가 두드러지게 크게 나타면서, 19차, 20차 당대회 보고문에서는 몇 개의 키워드로 집중되는 양상을 보인다. 이에 비해 18차 당대회 보고문에서의 키워드에서의 크기가 고른 분포를 보이는 것을 눈여겨 볼 필요가 있다.

3. 주요 표현 분석

이 절에서는 주요 표현(N-gram), 어휘의 조합을 살펴볼 것이다. 이를 통해 각 당대회 보고문의 성격이 보다 명확히 드러날 것이라 기대한다.

주요 표현을 추출하기 위해 AntConc의 N-gram 기능을 사용했다. N-gram을 추출할 때 크기 2, 최소 빈도수 1로 설정해, 두 단어로 이뤄진 표현 중 한 번 이상 사용된 모든 표현을 추출했다. 그 결과를 아래 [표 11]과 같이 정리했다.

www.kci.go.kr

[표 11] 당대회 보고문 중 N-gram(n=2) 추출 개요

차수	18차		19차		20차	
주요 표현	표현 종류	8446	표현 종류	9306	표현 종류	4222
	총 누적횟수	10544	총 누적횟수	11535	총 누적횟수	5077

N-gram(n=2)을 활용해 18차, 19차, 20차 당대회 보고문의 표현을 분석한 결과 전체 N-gram(n=2)의 종류와 총 누적횟수(토큰의 수)는 대체로 각 문서의 크기와 비례했다. 그 가운데 출현 빈도수 2 이상의 주요 표현은 문서의 크기와 상관없이 각각의 누적비율이 모두 25~30%로 수렴하는 경향을 보였다.

이 중 각 당대회의 빈도수 상위 N-gram(n=2)를 정리하면 [표 12]와 같다.

[표 12] 당대회 보고문 중 N-gram(n=2) 주요 표현

차수	18차		19차		20차	
구분	N-Gram(2)	freq	N-Gram(2)	freq	N-Gram(2)	freq
1	社会 管理	16	党 领导	21	国家 安全	17
2	经济 发展	16	实现 中华民族伟大复兴	19	全面建设社会主义 现代化	16
3	和平 发展	15	国家 安全	18	社会主义现代化 国家	16
4	发展 中国特色社会主义	13	中国人民	17	全面 推进	13
5	现代化 建设	13	党 建设	16	加快 建设	10
6	体制 改革	11	香港 澳门	15	新 发展	10
7	公共 服务	11	党 国家	14	党 国家	9
8	坚持 发展	11	美好 生活	14	全国 各族	9
9	科学 发展	11	体系 建设	13		

www.kci.go.kr

10	体制 机制	10	新时代 中国特色社会主义	13		
11	发展 方式	10	治理 体系	13		
12	服务 体系	10	体制 改革	12		
13	社会 发展	10	坚持 党	12		
14	社会 建设	10	中国特色社会主义 伟大	11		
15	经济 建设	10	人民 当家作主	11		
16	经济 社会	10	全体 人民	11		
17			发展 中国特色社会主义	11		
18			坚持 发展	11		
…						
175	国家 安全	4				

위의 결과를 보면, 주요 표현(N-gram(n=2))은 주로 당대회 보고문에서 빈도수 또는 핵심도가 높은 단어의 조합으로 수렴하면서 당대회의 핵심 안건이 어디에 있는지를 분명하게 보여준다. ‘发展’이 빈도수와 핵심도에서 1위를 차지했던 18차 당대회 보고문에서는 ‘经济+发展’(빈도수 16), ‘和平+发展’(빈도수 15), ‘发展+中国特色社会主义’(빈도수 13), ‘科学+发展’(빈도수 11) 등이 주요 표현으로 자주 쓰였다. ‘发展’이 빈도수 1위였으나 ‘党’이 핵심도 1위였던 19차 당대회 보고문에서는 ‘党+领导’(빈도수 21), ‘党+建设’(빈도수 16), ‘党+国家’(빈도수 14) 등이 주로 쓰이면서 당의 중심성을 강화했다. 흥미롭게도 20차 당대회 보고문의 주요 표현에서는 이런 빈도수 또는 핵심도로의 수렴 현상이 두드러지지 않는다. ‘国家+安全’이라는 표현이 빈도수 17로 가장 빈번하게 사용되었다.

이 결과를 보면 빈도수와 키워드 조사에서는 두드러지지 않다가 주요 표현 조사에서 두드러진 표현이 있다. 필자는 이 표현이 그 정권의

www.kci.go.kr

핵심 목표에 가까울 것이라 짐작한다. 18차 당대회 보고문에서는 '社会+管理'(빈도수 16)와 '公共+ 服务'(빈도수 11), 19차 당대회 보고문에서는 '国家+安全'(빈도수 18), '香港+澳门'(빈도수 15), '美好+生活'(빈도수 14), 20차 당대회 보고문에서는 압도적으로 '国家+安全'(빈도수 17)이다. 예를 들어 20차 당대회 보고문의 빈도수 1위 주요 표현 '国家+安全'은 18차 당대회 보고문에서는 빈도수 4로 175위였다가 19차 당대회 보고문에서 빈도수 18로 주요 표현 3위로 급상승했다. 19차 당대회 보고문에서 주요하게 쓰이다가 20차 당대회 보고문에서 가장 중심적으로 쓰이고 있는 '国家+安全'은 미중 대결 국면의 장기화와 코로나, 그리고 이로 인한 경제 위기 등 국내외 위기에 대한 대응으로 보인다.

다음으로 주요 표현들에 대한 의존도를 조사하기 위해 누적 빈도율을 조사, 정리했다.

[표 13] 18차 당대회 보고문 N-gram(n=2) 결과 누적 빈도율

18차						
번호	빈도수	누적빈도	비율(>=1)	누적(>=1)	비율(>=2)	누적(>=2)
1	16	16	0.15	0.15	0.51	0.51
2	16	32	0.15	0.30	0.51	1.01
3	15	47	0.14	0.45	0.47	1.49
4	13	60	0.12	0.57	0.41	1.90
5	13	73	0.12	0.69	0.41	2.31
6	11	84	0.10	0.80	0.35	2.66
7	11	95	0.10	0.90	0.35	3.00
8	11	106	0.10	1.01	0.35	3.35
9	11	117	0.10	1.11	0.35	3.70
10	10	127	0.09	1.20	0.32	4.02
11	10	137	0.09	1.30	0.32	4.33
12	10	147	0.09	1.39	0.32	4.65
13	10	157	0.09	1.49	0.32	4.96
14	10	167	0.09	1.58	0.32	5.28
15	10	177	0.09	1.68	0.32	5.60
16	10	187	0.09	1.77	0.32	5.91

7

www.kci.go.kr

17	9	196	0.09	1.86	0.28	6.20	
18	9	205	0.09	1.94	0.28	6.48	11
19	9	214	0.09	2.03	0.28	6.77	
20	9	223	0.09	2.11	0.28	7.05	
21	9	232	0.09	2.20	0.28	7.33	
22	9	241	0.09	2.29	0.28	7.62	
23	9	250	0.09	2.37	0.28	7.90	
24	8	258	0.08	2.45	0.25	8.16	
25	8	266	0.08	2.52	0.25	8.41	
26	8	274	0.08	2.60	0.25	8.66	
27	8	282	0.08	2.67	0.25	8.92	
28	8	290	0.08	2.75	0.25	9.17	
29	8	298	0.08	2.83	0.25	9.42	
30	8	306	0.08	2.90	0.25	9.67	
31	8	314	0.08	2.98	0.25	9.93	13
32	8	322	0.08	3.05	0.25	10.18	
33	8	330	0.08	3.13	0.25	10.43	
34	8	338	0.08	3.21	0.25	10.69	
35	8	346	0.08	3.28	0.25	10.94	
36	8	354	0.08	3.36	0.25	11.19	

[표 14] 19차 당대회 보고문 N-gram(n=2) 결과 누적 빈도율

19차							
번호	빈도수	누적빈도	비율(>=1)	누적(>=1)	비율(>=2)	누적(>=2)	
1	21	21	0.18	0.18	0.63	0.63	
2	19	40	0.16	0.35	0.57	1.20	
3	18	58	0.16	0.50	0.54	1.74	
4	17	75	0.15	0.65	0.51	2.25	
5	16	91	0.14	0.79	0.48	2.73	
6	15	106	0.13	0.92	0.45	3.18	6
7	14	120	0.12	1.04	0.42	3.59	
8	14	134	0.12	1.16	0.42	4.01	
9	13	147	0.11	1.27	0.39	4.40	
10	13	160	0.11	1.39	0.39	4.79	
11	13	173	0.11	1.50	0.39	5.18	
12	12	185	0.10	1.60	0.36	5.54	

www.kci.go.kr

13	12	197	0.10	1.71	0.36	5.90	
14	11	208	0.10	1.80	0.33	6.23	
15	11	219	0.10	1.90	0.33	6.56	
16	11	230	0.10	1.99	0.33	6.89	10
17	11	241	0.10	2.09	0.33	7.22	
18	11	252	0.10	2.18	0.33	7.55	
19	10	262	0.09	2.27	0.30	7.85	
20	10	272	0.09	2.36	0.30	8.15	
21	10	282	0.09	2.44	0.30	8.45	
22	10	292	0.09	2.53	0.30	8.75	
23	10	302	0.09	2.62	0.30	9.05	
24	10	312	0.09	2.70	0.30	9.35	
25	10	322	0.09	2.79	0.30	9.65	
26	10	332	0.09	2.88	0.30	9.95	
27	10	342	0.09	2.96	0.30	10.25	11
28	9	351	0.08	3.04	0.27	10.52	
29	9	360	0.08	3.12	0.27	10.78	
30	9	369	0.08	3.20	0.27	11.05	
31	9	378	0.08	3.28	0.27	11.32	
32	9	387	0.08	3.36	0.27	11.59	

[표 15] 20차 당대회 보고문 N-gram(n=2) 결과 누적 빈도율

20차							
번호	**빈도수**	**누적빈도**	**비율(>=1)**	**누적(>=1)**	**비율(>=2)**	**누적(>=2)**	
1	17	17	0.33	0.33	1.28	1.28	
2	16	33	0.32	0.65	1.20	2.48	
3	16	49	0.32	0.97	1.20	3.69	3
4	13	62	0.26	1.22	0.98	4.67	
5	10	72	0.20	1.42	0.75	5.42	
6	10	82	0.20	1.62	0.75	6.17	
7	9	91	0.18	1.79	0.68	6.85	
8	9	100	0.18	1.97	0.68	7.52	5
9	9	109	0.18	2.15	0.68	8.20	

www.kci.go.kr

10	8	117	0.16	2.30	0.60	8.80	
11	8	125	0.16	2.46	0.60	9.41	
12	8	133	0.16	2.62	0.60	10.01	
13	8	141	0.16	2.78	0.60	10.61	
14	8	149	0.16	2.93	0.60	11.21	6
15	8	157	0.16	3.09	0.60	11.81	

[표 13], [표 14], [표 15]에서 보이듯, 주요 표현들의 집중도에 있어서는 20차 당대회 보고문이 다른 양상을 드러냈다. 20차 당대회 보고문의 경우 빈도수 최상위 3개 표현의 누적비율이 1%에 육박함으로써 18차와 19차 보고문에서 같은 수준에 도달하기까지 각기 7개와 6개의 표현이 필요했던 것과는 유의미한 차이를 보였다. 특히 20차 보고문의 빈도수 1,2위 표현인 '国家+安全'과 '全面建设+社会主义现代化'는 전체 N-gram 토큰에서 차지하는 비율이 각기 0.33%와 0.32%로 18차 및 19차 빈도수 1,2위 표현 대비 출현비율이 2배 이상 높은 것은 주목할 만한 현상이라고 하겠다. 20차 보고문에서 빈도수 17, 출현비율 0.33%로 1위 표현인 '国家+安全'은 18차 보고문의 경우 빈도수 4, 출현비율 0.04%에 불과했으며, 19차에서는 주요 표현 4위로 급상승하며 빈도수 18을 기록했지만 출현비율은 0.16% 정도였다. 따라서 20차에서 1위 주요 표현으로 올라선 '国家+安全'에 대한 중요도는 19차 대비 2배 이상 상승했다고 말할 수 있으며, 그 비중은 18차 및 20차의 1위 주요 표현인 '社会+管理'(빈도수 16, 출현비율 0.15%)와 '党+领导'(빈도수 21, 출현비율 0.18%)에 비해서도 매우 높은 것이라 할 수 있다.

Ⅳ. 결론 및 향후 과제

지금까지 텍스트 마이닝 방법론을 이용해 20차 당대회 보고문의 어

www.kci.go.kr

휘 빈도 분석, 키워드 분석, 주요 표현 분석을 진행했다. 또한 이 결과를 18차, 19차 당대회 보고문에 대한 분석 결과와 비교함으로써 20차 당대회 보고문의 특징을 드러내고자 시도했고 이를 통해 몇 가지 특징을 도출할 수 있었다.

첫째, 어휘 빈도 분석에서 18차, 19차, 20차 당대회 보고문 공히 '发展'이 가장 높은 빈도수를 보였다. 그러나 이 '发展'이 앞에 다른 어휘와 결합해 주요 표현으로 등장할 때 20차 당대회 보고문의 경우 '新+发展', '高质量+发展' 등의 새로운 개념어로 사용된 반면, 18차에서는 '经济+发展', '科学+发展' 등 명사와 결합하여 발전의 영역을, 19차에서는 '坚持+发展', '和平+发展' 등 동사 및 형용사와 결합하여 발전의 양상을 강조하는 방향으로 사용되었다.

둘째, 키워드 분석에서는 20차 당대회 보고문의 경우 '坚持', 19차의 경우 '党', 18차의 경우 '发展'이 핵심도 1위의 키워드로 나타났다. 18차, 19차, 20차 당대회 보고문 공히 빈도수 1위 어휘가 '发展'이었음에도 핵심도가 반영된 키워드의 순위에서 변화가 나타난 것은 20차에서 '坚持', 19차에서 '党', 18차에서 '发展'이 참조 코퍼스 대비 상대적으로 가장 중요하게 사용된 어휘임을 의미한다. 키워드 분석에서 18차는 핵심도 값이 비교적 고른 분포를 보이지만 19차와 20차는 상위 몇 개 키워드의 핵심도가 상대적으로 높게 나타났다. 이는 18차에 비해 19차, 20차 당대회 보고문의 상위 키워드에 대한 의존도가 상대적으로 높은 것이라 볼 수 있다.

셋째, 주요 표현(N-gram(n=2)) 분석에서는 20차 당대회 보고문의 경우 '国家+安全', 19차의 경우 '党+领导', 18차의 경우 '社会+管理'와 '经济+发展'이 나란히 빈도수 1위로 나타났다. 20차의 주요 표현(N-gram(n=2)) 빈도수 1위인 '国家+安全'은 18차에서 174위였다가 19차에서 3위로 뛰어올라 20차에서 1위가 된 표현이다. 또한 주요 표현의 집중도에 있어서도 20차 당대회 보고문의 상위 표현들이 상대적으로

www.kci.go.kr

높게 나타났다.

넷째, 이상의 어휘 빈도, 키워드, 주요 표현의 분석을 워드클라우드로 시각화해 살펴본 결과 각기 상위에 랭킹된 어휘들에 대한 텍스트의 의존 내지 집중도는 20차 당대회 보고문이 상대적으로 가장 높고, 18차 당대회 보고문이 가장 균형적인 고른 분포를 나타내는 것으로 드러났다. 결론적으로 20차 당대회 보고문이 어휘를 통한 의미 구성에 있어 18차 및 19차에 비해 상대적으로 집중적인 수렴형 양상을 보이고 18차 당대회 보고문이 상대적으로 균형적인 확산형 양상을 보인다고 해석할 수 있다.

지금까지 연구를 진행하며 기존에 당대회 문건을 텍스트로 독해할 때 미처 발견하지 못했던 점을 계량화, 시각화할 수 있게 되었다. 특히 각각의 당대회 보고문에서 공히 빈도수 1위인 '发展'이라는 어휘가 서로 다른 말과 결합해 서로 다른 맥락을 형성하는 점을 보여주었는데, 이를 다른 차수 당대회 보고문으로 확대해 비교하면 더욱 흥미로운 결과를 얻을 수 있을 것이다. 또한 본 연구는 키워드 분석에서 15가지 다양한 문체를 참조 코퍼스로 삼았는데, 이를 더욱 확대하거나 이전 당대회 보고문으로 설정하는 등 다양한 시도를 해본다면 더욱 의미 있는 결과를 얻을 수 있을 것이다.

마지막으로 전체 연구 과정에서 여러 프로그램을 이용해 최대한 정밀하고 정확하게 진행하려 노력했다. 그럼에도 불구하고 많은 단계에서 연구자의 깊이 있는 판단과 전문 역량이 필요한 부분이 많았다. 연구 분석에 쓰이는 데이터의 신뢰도를 높이고 투명하게 만드는 것은 향후 계속 풀어야 할 숙제가 될 것이다.

www.kci.go.kr

참고문헌

중국공산당원 공식 홈페이지(https://www.12371.cn/)

박소희.(2022). 제20차 당 대회로 본 중국의 미중 전략경쟁 대응 전략. *월간 KIET 산업경제* vol.290.

양갑용.(2022). 중국공산당 20차 당대회와 '70후'. *성균차이나브리프*.

주장환.(2022). 중국 엘리트 정치 동학 변화에 관한 연구: 제20차 공산당 전국대표대회와 제20기 중앙위원회 1차 전체회의를 중심으로, *21세기정치학회보* vol. 32.

성균중국연구소.(2022). 중국공산당 제20차 전국대표대회 – 중국식 현대화 추진을 위한 시진핑 친정체제 구축, *SICS연구보고서 22-03*.

표나리.(2022). 중국공산당 제20차 당대회 분석: 주요 쟁점과 외교적 함의, *주요국제문제분석* 2022-34.

Heather Froehlich.(2015). Corpus Analysis with Antconc. *Programming Historian.* https://programminghistorian.org/en/lessons/corpus-analysis-with-antconc

Laurence Anthony.(2012). *AntConc(Windows, Macintosh OS X, and Linux)*.https://www.laurenceanthony.net/software/antconc/releases/AntConc335/help.pdf

Manuscript: Jan 22, 2023; Review completed: Jan 28, 2023; Accepted: Feb 03, 2023

www.kci.go.kr