

중국지도자 신년사에 대한 양적 비교분석*

— 江澤民, 胡錦濤, 習近平 간 TTR, TF-IDF 비교를 중심으로 —

박 찬 욱**

<目次>

- | | |
|-----------------|---------------------|
| I. 서론 | IV. 신년사 핵심어와 시대 키워드 |
| II. 언어자료와 분석도구 | 1. 신년사의 핵심어 비교 |
| III. 신년사의 어휘다양도 | 2. 신년사와 시대 키워드 |
| 1. 빈도 추이와 TTR | 3. 신년사와 키워드 주제 |
| 2. 고빈도 어휘 비교 | V. 결론 |

I. 서론

본고는 1993년부터 2022년까지 발표된 장쩌민(江澤民), 후진타오(胡錦濤), 시진핑(習近平)의 신년사를 양적으로 비교 분석함으로써 지도자의 언어가 시대 맥락을 어떻게 담아내고 있는지 고찰하는 데 목적이 있다.¹⁾

중국은 매해 12월 말이면 재임 중인 국가 주석이 중앙매체(中國中央電視臺, CCTV)를 통해 대국민 신년 담화를 발표한다. 국가주석이 정부와

* 본고는 2023년 5월 20일, 부경대학교에서 개최된 ‘大韓中國學會 創立40周年紀念 2023年度 春季學術大會’에서 ‘중국 지도자의 신년사에 나타난 핵심어 비교 분석 — 1993년~2022년 신년사에 대한 양적분석을 중심으로’라는 제하로 구두 발표한 내용을 수정, 보완하였다.

** 경희대학교 중국어학과 교수

1) 신년사는 12월 31일 CCTV를 통해 담화로 발표되고 이듬해 1월 1일 《人民日報》에 실린다. 본고의 신년사 ‘발표’년도는 방송매체를 기준으로 한다.

당을 대표하여 약 10분 내외로 대국민 담화를 발표하는데 ‘新年賀詞’라는 제하의 담화임에도 그 내용은 ‘賀詞’에 국한하지 않는다. 지도자(그룹)가 주창한 사상과 이념, 정치와 경제, 기술과 교육, 민생과 복지 등에 걸쳐 당해 연도의 연간 성과와 향후 비전 및 계획 등을 담고 있다는 점에서, 신년사는 단순 하례를 넘어선 공적 정치담화로 간주될 수 있다.²⁾

중국의 지도자 담화에 대한 연구로는, 장쩌민의 신년사를 사회적 행위 구조와 我們 간 관계의 관점에서 분석한 박찬욱(2009), 시진핑 講話에서의 我們을 비판적 담화분석의 관점에서 접근한 이지원(2020), 시진핑 講話와 후진타오 文選을 텍스트마이닝 방법으로써 비교 분석한 이경진 등(2023)을 들 수 있다. 이들 연구는 개별 지도자 담화에 상대적으로 치중해 있고 이경진 등(2023)을 제외하면 기술통계 측면에서만 양적 접근을 해왔다는 한계도 있다. 무엇보다 문제로서의 신년사는 박찬욱(2009), Hope(2025) 등 몇몇 논의로부터 알 수 있듯 국내외적으로 충분한 관심을 받지 못했다. 생각건대, 단일 편폭으로는 데이터가 충분하지 않다는 점, 이듬해 신년사를 수집하기 위해서는 다시 일 년이 소요되는 점, 그리고 전국인민대표대회나 전국대표대회처럼 정치담화의 상황적 맥락이 다소 결여되어 있다는 점 등을 그 이유로 꼽을 수 있을 것이다.

2022년 기준으로, 장쩌민, 후진타오, 시진핑으로 이어진 재임 기간의 총합은 30년이다. 그에 따라 신년사도 매해 한 회씩 총 30회, 약 300분의 분량이 쌓였다. 30년의 누적된 시간량 속에서 출현한 어휘와 시대 간 연관성에 대한 양적 접근은 중국을 사회언어적 관점에서 이해하는 데 도움을 줄 것이다. 더불어 장쩌민의 10년, 후진타오의 10년, 시진핑의 10년이라는 시간량은 각 지도자들의 연도별, 임기별 공통점과 차이점을 살피는데 적절한 균형자 역할을 할 수 있을 것이다.

이하 본론에서는 먼저 언어자료 개요와 연구에 사용된 도구를 간략히 소개한 뒤, 문장과 어휘 빈도 및 불용어 제거 후 산출된 내용어 비율을 통

2) 박찬욱, 〈중국국가지도자 신년사의 행위구조와 ‘我們’ — 江澤민의 1997년, 1999년 신년사를 중심으로〉, 《중국어문학논집》, 제57호, 2009, 166쪽.

해 어휘다양도(TTR)를 살펴본다. 이어서, 고빈도 어휘 100개를 추출하여 지도자 3인의 고빈어 어휘사용 양상을 관찰한다. 그리고 어휘빈도와 역문서빈도 간 연산(TF-IDF)에 기초하여 신년사 속 핵심어를 가늠하고 신년사와 시대별 키워드 간 연관성 등을 분석한다. 끝으로, 논의 내용을 정리하고 함축을 제시한다.

II. 언어자료와 분석도구

1. 언어자료

본고는 ‘人民日報’³⁾, ‘人民網’⁴⁾, ‘大公網’⁵⁾, ‘百度百科’⁶⁾ 등으로부터 언어자료 42,817자(.txt 기준)를 수집하였다. 웹이라는 공간에서 필요한 자료 수집이 가능해진 만큼 언어자료의 수집 자체는 과거에 비해 상대적으로 용이해졌다. 다만 수집된 자료 규모가 커지면서 언어자료의 관리와 활용에도 더 많은 주의를 요구하고 있다. 이에 본고는 텍스트파일(.txt)로 자료를 수집하고 공란과 단락구분의 소거, 문장부호와 간체-번체, 숫자 byte 통일 등의 정제를 한 뒤 파이썬을 이용하여 JSON(.json) 형식으로 변환하였다. 이를 통해 판다스(pandas) 등의 파이썬 라이브러리나 SQLite 같은 데이터베이스 쿼리(query)로도 접근이 가능하도록 하였다. 언어자료를 전처리하는 데 도움을 주는 파이썬 라이브러리는 다양하게 존재하는데, 본고는 문

3) 人民日報: <https://reader.jojokanbao.cn/rmrh>.

4) 人民網(習近平系列重要講話數據庫): <http://jhsjk.people.cn/>.

5) 大公網(元旦賀詞): <http://news.takungpao.com/special/yuandanheci/>.

6) 百度百科(1992年-2025年新年賀詞): https://baike.baidu.com/starmap/view?noDeId=dfe1324092db5f68bfb97f8b&lemmaTitle=1997%E5%B9%B4%E6%96%B0%E5%B9%B4%E8%B4%BA%E8%AF%8D&lemmaId=24241066&starMapFrom=lemma_starMap&fromModule=lemma_starMap.

장분리(sentence segmentation)에서 Stanza(ver.1.11.0)⁷⁾를 사용하고 단어분리(word segmentation)와 품사태깅(pos tagging)에서 HanLP(ver.2.1.3)⁸⁾를 사용하였다.

2. 분석도구

본고는 상기 JSON 변환부터 정량화 작업까지 바이브코딩(Vibe Coding) 방법을 이용하였다. 바이브코딩이란 “직감(Vibe)과 느낌에 의존하여 AI의 도움을 받아 코드를 작성하는 개발 방식”을 의미하는데,⁹⁾ 본고에서는, 구체적으로, 연구계획과 문제의식, 연구주제와 예상 출력형식을 프롬프트로 설명하고 결과 구현에 필요한 코드와 관련 로직을 AI로부터 지원받는 방식으로 진행하였다. 이는 자연어에 의지한 코딩으로써 연구자가 코딩에 대한 기술적 숙련보다 연구 목적과 아이디어 구현에 좀 더 집중할 수 있도록 하는 장점을 갖는다. 이 같은 취지하에 본고는 구글의 생성형 AI인 Gemini(ver. 2.5 flash, Canvas)¹⁰⁾를 이용하였고, AI에게 역할을 부여한 뒤 맥락과 예시, 출력 형식을 제시하며 단계적인 사고로써 응답하도록 프롬프트를 제공하였다. 프롬프트의 예는 <표 1>과 같다.

<표 1> 정량 작업을 위한 코딩 개시 프롬프트

>> 너는 중국어를 정량 분석하는 전문가야.
 >> 다음 연구주제는 * *야. 내가 생각하는 **○○○란 **를 가리켜. 그래서 ○○방법론을 이용하여 를 분석할 생각이야. 작업에 앞서 아래 <주의사항>이 있으니 기억해야해.
 <주의사항>
 1. 분석 대상은 오직 “ ” 뿐이야. 더미 데이터를 생성하면 안 돼.

 >> 나의 <문제의식>은 아래와 같아.

7) Stanza: <https://stanfordnlp.github.io/stanza/>.

8) HanLP: <https://pypi.org/project/hanlp/>.

9) 오힘찬, 《제미나이 노트북LM 미친활용법 51제》, 서울: 골든래빗, 2025, 296쪽.

10) Gemini: <https://gemini.google.com/app>.

```

<문제의식>
1. ○○방법론으로 중국의 지도자들의 ..... 를 관찰할 수 있을까
.....
>> <문제의식>에 따라 분석하려는 내용은 아래와 같아.
<분석내용>
> 분석주제: .....
    1. ○○를 이용하여 추출된 .....
.....
>> *분석주제*에서의 “○○”란 “.....”를 가리켜. 그래서 다음과 같이 <키워드>를 제시하니 참고해.
<키워드>
.....
>> ..... 관련 데이터를 파악할 수 있도록 <주의사항> 다시 한 번 숙지한 뒤
>> <분석내용>을 단계적으로 생각한 후 <문제의식>에 타게팅 된 코드를 작성해줘.
>> 코드 실행 결과는 아래의 출력 포맷을 가져야 해.
<format>
1. .... 핵심어 :
2. .... 수치 :
    
```

소기한 형식의 출력값을 내기까지 지난한 ‘질문—대답’이 이어지는데, 이 과정에서 가장 빈번하게 부딪히는 오류로는 (AI가 연구자의 작업환경을 알지 못해서 발생하는) 입력데이터 오류를 꼽을 수 있다. 또 적용할 데이터를 적절하게 제공하지 않고 코드만 요청할 경우, 생성형 AI는 자신의 플랫폼에서 가장 훌륭한 결과값을 내보이기 위해 시뮬레이션용 데이터를 임의 생산하여 대답하기도 했다. 그럼에도 하나의 대화 흐름을 유지하면서 이전 기록을 지속해 소환함으로써 코드를 수정하고 재현성을 가질 수 있도록 여러 생성형 AI를 사용하는 것은 지양하였다.

Ⅲ. 신년사의 어휘다양도

1. 빈도 추이와 TTR

먼저 30년 신년사에 보이는 문장, 어휘 빈도값을 살펴본다. 결과값은 아

래와 같다.

- (1) 분석된 연설 수 (총 문서 수): 30개
 · 분석된 레코드 수 (총 문장 수): 1036개
 1. 평균 문장수 (연설 당): 35개
 2. 평균 어휘수 (연설 당): 882개
 3. 평균 내용어수 (연설 당, 불용어 제외): 531개

총 30회 연설에서 문장은 총 1036개, 연설 당 평균 문장수는 35개이다. 연설 당 평균 사용자어휘수는 882개인데 이 수치에는 기능어와 같이 “많은 정보를 전달하지 않지만 출현 빈도가 높기 때문에 중요한 내용을 놓치게” 하는 불용어(stopword)가 포함되어 있다.¹¹⁾ 불용어에 정확한 기준은 없으나 본고는 범용돼온 4개(百度停用詞表, 中文停用詞表, 哈工大停用詞表, 四川大學機器智能實驗室停用詞庫) 중에서 ‘中文停用詞表’를 이용하여 처리하였다.¹²⁾ 이에 따라 불용어처리 후 연설 당 평균 내용어수는 531개가 산출되었다. 이에 관한 상세는 아래와 같다.

〈표 2〉 중국지도자 신년사 문장빈도, 어휘빈도, TTR

지도자	총연설수	총어휘수	총고유어휘수	총내용어수	총고유내용어수	TTR (Type-Token Ratio)	평균문장수 (연설당)	평균어휘수 (연설당)	평균내용어수 (연설당)
ALL	30	26,456	3,349	15,936	3,124	0.196	35	882	531
JIANG	10	9,735	1,479	5,807	1,303	0.224	40	974	581
HU	10	7,344	950	4,680	843	0.180	22	734	468
XI	10	9,377	2,386	5,449	2,217	0.407	42	938	545

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑

30년 동안 지도자 신년사에 사용된 ‘총어휘수’는 26,456개이고 이 중에서

11) 젠스 알브레히트, 싯다르트 라마찬드란, 크리스티안 뵘클러 저, 심상진 역, 《과이썬 라이브러리를 활용한 텍스트 분석》, 서울: 한빛미디어, 2022, 43쪽.

12) GitHub(goto456/stopwords) : <https://github.com/goto456/stopwords>.

중복을 제외한 ‘총고유어휘수’는 3,349개이다. 총어휘수에서 불용어를 제외한 ‘총내용어수’는 15,936개이고 이 중 중복을 제거한 ‘총고유내용어수’는 3,124개이다. 어휘다양도로 불리는 TTR(Type-Token Ratio)은 텍스트에서 어휘가 얼마나 다양하게 출현하는가를 측정하는 지표로서 흔히 교재 난이도 측정에 사용되며 총고유어휘수(Type)를 총어휘수(Token)로 나누어 산출한다.¹³⁾ 다만 본고는 난이도 측정이 아닌, 얼마나 다양한 내용어로 실질 메시지를 전달하는가에 측정 목적을 두기 때문에 기존의 ‘Type’을 ‘총내용어수’로, ‘Token’을 ‘총고유내용어수’로 조정하여 산출한다. 이에 본고의 TTR은 어휘다양도, 그 중에서도 고유내용어의 다양도를 가리킨다. TTR은 다음의 수식으로 표현되며 1에 가까울수록 여러 어휘를 다양하게 사용한 것으로, 낮으면 동일한 어휘를 반복하여 사용한 것으로 판단한다.

$$(2) TTR = \frac{\text{총고유내용어수}}{\text{총내용어수}}$$

이로써 산출된 지도자 전체 TTR은 0.196이다. 신년사에서는 세 지도자가 동일 어휘를 반복 사용했을 가능성이 상대적으로 높다고 할 수 있는데, 이는 지도자 신년사라는 특정 포맷하에서, 다양한 어휘보다는 한정된 어휘를 반복 사용함으로써 일반 대중들에게 정부와 당의 대국민 메시지를 일관되게 전달하려는 전략에 기인한다.

다음으로 지도자 개인의 관점에서 살펴보자. ‘총어휘수’는 장쩌민이 9,735개로 가장 많고 시진핑 9,377개, 후진타오 7,344개로 뒤를 잇는다. 그런데 중복을 제외한 ‘총고유어휘수’는 시진핑이 2,386개로 가장 많고 장쩌민 1,479개, 후진타오 950개 순이다. 불용어를 제외한 ‘총내용어수’는 장쩌민이 5,807개로 가장 많고 시진핑 5,449개, 후진타오 4,680개 순이다. 그런데 불용어와 중복을 모두 제거한 ‘총고유내용어수’에서는 다시 시진핑

13) 김가현, 김지영, <절대평가 도입 전·후 수험능력시험 영어 읽기 지문의 어휘 분석 비교>, 《외국학연구》, 제58집, 2021, 14쪽.

2,217개, 장쩌민 1,303개, 후진타오 843개 순으로써 ‘총어휘수’ 대비 ‘총고유어휘수’와 패턴을 같이 하고 있다.

이 같은 순서와 패턴은 TTR에 반영되어 시진핑 0.407, 장쩌민 0.224, 후진타오 0.180 순을 보였다. 시진핑은 여러 어휘를 다양하게 사용하며 메시지를 전달하는 데 반해, 후진타오는 시진핑의 절반 이하 수치로서 동일한 어휘를 반복 사용하는 경향이 상대적으로 높다고 할 수 있다. 장쩌민 역시, 총어휘수와 총내용어수가 많음에도 불구하고 중복을 제외하면 어휘수가 큰 폭으로 감소하면서 시진핑보다는 후진타오에 TTR이 가까워지며 메시지 전달에 있어 어휘를 중복사용하는 경향을 보였다.

평균 문장수는 시진핑 42개, 장쩌민 40개, 후진타오 22개 순으로 나타났다. TTR과 함께 문장수에서도 시진핑이 장쩌민에 앞서 있는 결과이다. 이는, 의사소통의 최소단위가 문장으로 간주된다는 점을 고려하면, 시진핑의 신년사가 장쩌민, 후진타오의 그것보다 상대적으로 다양한 어휘와 많은 정보량으로 구성되어 있음을 시사한다.

2. 고빈도 어휘 비교

다음으로, 지도자 신년사의 고빈도 어휘 상위 100개를 살펴본다. 우선 수치 간 합산 결과이다.

〈표 3〉 중국지도자 신년사 고빈도 어휘 상황

지도자	총어휘수	총내용어수	상위 100개 단어빈도	총어휘수 대비 비율	총내용어수 대비 비율
ALL	26,102	15,798	6,814	26.11%	43.13%
JIANG	9,653	5,770	2,891	29.95%	50.10%
HU	7,214	4,639	2,625	36.39%	56.59%
XI	9,235	5,389	1,758	19.04%	32.62%

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑

〈표 3〉에서 관찰된 몇 가지 패턴을 살펴보자. 첫째, 후진타오는 ‘총어휘

수'에서 장쩌민, 시진핑보다 약 2,000여 개 이상 적음에도 불구하고 상위 100개 어휘가 전체 신년사의 36.39%(불용어처리 전), 56.59%(불용어처리 후)를 점하고 있다. 메시지 전달에 절반이 사용될 정도로 상위 100개 어휘의 사용이 두드러진다. 둘째, 후진타오와 마찬가지로, 장쩌민도 29.95%(불용어처리 전), 50.10%(불용어처리 후)로서 메시지 전달에 상위 100개 어휘의 역할이 두드러진다. 그에 반해, 시진핑은 각각 19.04%(불용어처리 전), 32.62%(불용어처리 후)로서, 전임 두 지도자들과는 다르게, 약 70% 가까운 어휘를 상대적으로 낮은 빈도와 비율로 다양하게 사용하고 있다. 다음은 지도자들의 상위 100개 어휘목록이다.

〈표 4〉 중국지도자 신년사 상위 100개 어휘 목록

순위	ALL	F	T(%)	C(%)	JIANG	F	T(%)	C(%)	HU	F	T(%)	C(%)	XI	F	T(%)	C(%)
1	中國	459	1.7855	2.9054	中國	228	2.3620	3.9515	發展	141	1.9545	3.0394	中國	105	1.137	1.9484
2	人民	350	1.3409	2.2155	人民	143	1.4814	2.4783	中國	126	1.7466	2.7161	人民	87	0.9421	1.6144
3	發展	322	1.2336	2.0382	發展	136	1.4089	2.357	人民	120	1.6234	2.5868	年	69	0.7472	1.2804
4	國	278	1.0651	1.7597	世界	116	1.2017	2.0104	和平	103	1.4278	2.2203	國	67	0.7255	1.2433
5	世界	263	1.0076	1.6648	國	112	1.1603	1.9411	國	99	1.3723	2.1341	世界	48	0.5198	0.8907
6	和平	206	0.7892	1.3040	和平	86	0.8909	1.4905	世界	99	1.3723	2.1341	新	47	0.5089	0.8721
7	年	188	0.7203	1.1900	新	81	0.8391	1.4038	經濟	77	1.0674	1.6598	發展	45	0.4873	0.835
8	新	183	0.7011	1.1584	年	79	0.8184	1.3692	國際	62	0.8594	1.3365	改革	39	0.4223	0.7237
9	經濟	156	0.5977	0.9875	經濟	68	0.7044	1.1785	新	55	0.7624	1.1856	全面	34	0.3682	0.6309
10	國際	133	0.5095	0.8419	國際	54	0.5594	0.9359	合作	50	0.6931	1.0778	同胞	26	0.2815	0.4825
11	共同	109	0.4176	0.6900	建設	45	0.4662	0.7799	共同	47	0.6515	1.0131	全	25	0.2707	0.4639
12	繼續	108	0.4138	0.6836	繼續	44	0.4558	0.7626	同胞	45	0.6238	0.9700	國家	25	0.2707	0.4639
13	同胞	101	0.3869	0.6393	世紀	41	0.4247	0.7106	繼續	45	0.6238	0.9700	實現	25	0.2707	0.4639
14	社會	100	0.3831	0.6330	共同	39	0.4040	0.6759	兩	44	0.6099	0.9485	鐫刻	25	0.2707	0.4639
15	兩	98	0.3755	0.6203	兩	39	0.4040	0.6759	社會	43	0.5961	0.9269	新年	24	0.2599	0.4454
16	建設	97	0.3716	0.6140	中	38	0.3937	0.6586	年	40	0.5545	0.8623	共同	23	0.2491	0.4268
17	改革	94	0.3601	0.5950	社會主義	36	0.3729	0.6239	促進	37	0.5129	0.7976	生活	23	0.2491	0.4268
18	合作	86	0.3295	0.5444	社會	35	0.3626	0.6066	建設	37	0.5129	0.7976	群眾	23	0.2491	0.4268
19	香港	81	0.3103	0.5127	人類	35	0.3626	0.6066	香港	34	0.4713	0.7329	社會	22	0.2382	0.4082
20	促進	70	0.2682	0.4431	國家	34	0.3522	0.5893	推進	34	0.4713	0.7329	努力	20	0.2166	0.3711
21	全面	70	0.2682	0.4431	改革	33	0.3419	0.5719	堅持	34	0.4713	0.7329	美好	19	0.2057	0.3526
22	國家	69	0.2643	0.4368	民族	32	0.3315	0.5546	澳門	32	0.4436	0.6898	奮鬥	19	0.2057	0.3526
23	人類	68	0.2605	0.4304	臺灣	31	0.3211	0.5373	特別	28	0.3881	0.6036	繼續	19	0.2057	0.3526
24	社會主義	67	0.2567	0.4241	同胞	30	0.3108	0.5199	行政區	28	0.3881	0.6036	黨	19	0.2057	0.3526
25	實現	67	0.2567	0.4241	文明	30	0.3108	0.5199	開放	28	0.3881	0.6036	香港	18	0.1949	0.334
26	新年	66	0.2529	0.4178	香港	29	0.3004	0.5026	社會主義	25	0.3465	0.5389	偉大	18	0.1949	0.334
27	努力	66	0.2529	0.4178	實現	28	0.2901	0.4853	全面	25	0.3465	0.5389	推進	17	0.1841	0.3155
28	中	65	0.2490	0.4114	合作	28	0.2901	0.4853	努力	25	0.3465	0.5389	民族	17	0.1841	0.3155
29	開放	65	0.2490	0.4114	促進	28	0.2901	0.4853	族	24	0.3327	0.5174	和平	17	0.1841	0.3155
30	推進	64	0.2452	0.4051	統一	26	0.2693	0.4506	問題	24	0.3327	0.5174	國際	17	0.1841	0.3155
31	全	62	0.2375	0.3925	里	25	0.2590	0.4333	維護	24	0.3327	0.5174	祝福	16	0.1733	0.2969
32	澳門	62	0.2375	0.3925	全	25	0.2590	0.4333	穩定	24	0.3327	0.5174	繼續	16	0.1733	0.2969
33	堅持	59	0.2260	0.3735	穩定	25	0.2590	0.4333	中央	22	0.3050	0.4742	中	16	0.1733	0.2969
34	生活	57	0.2184	0.3608	政治	24	0.2486	0.4159	改革	22	0.3050	0.4742	週年	16	0.1733	0.2969

10 中國語文學 第100輯

순위	ALL	F	T(%)	C(%)	JLJANG	F	T(%)	C(%)	HU	F	T(%)	C(%)	XI	F	T(%)	C(%)
35	穩定	55	0.2107	0.3481	進步	24	0.2486	0.4159	保持	22	0.3050	0.4742	希望	15	0.1624	0.2783
36	美好	55	0.2107	0.3481	祖國	24	0.2486	0.4159	方針	22	0.3050	0.4742	更加	15	0.1624	0.2783
37	民族	54	0.2069	0.3418	歷史	24	0.2486	0.4159	新年	21	0.2911	0.4527	開放	15	0.1624	0.2783
38	臺灣	52	0.1992	0.3292	朋友們	23	0.2383	0.3986	繁榮	21	0.2911	0.4527	建設	15	0.1624	0.2783
39	特別	52	0.1992	0.3292	開放	22	0.2279	0.3813	岸	21	0.2911	0.4527	北京	15	0.1624	0.2783
40	行政區	49	0.1877	0.3102	中華	22	0.2279	0.3813	廣播	20	0.2772	0.4311	兩	15	0.1624	0.2783
41	朋友們	48	0.1839	0.3038	新年	21	0.2175	0.3640	電臺	20	0.2772	0.4311	號	15	0.1624	0.2783
42	里	48	0.1839	0.3038	取得	21	0.2175	0.3640	治	20	0.2772	0.4311	會	14	0.1516	0.2598
43	問題	47	0.1801	0.2975	努力	21	0.2175	0.3640	加融	20	0.2772	0.4311	人口	14	0.1516	0.2598
44	祖國	46	0.1762	0.2912	加融	19	0.1968	0.3293	人類	20	0.2772	0.4311	特別	13	0.1408	0.2412
45	繁榮	45	0.1724	0.2848	重要	19	0.1968	0.3293	美好	19	0.2634	0.4096	創造	13	0.1408	0.2412
46	積極	45	0.1724	0.2848	澳門	19	0.1968	0.3293	積極	19	0.2634	0.4096	必	13	0.1408	0.2412
47	統一	45	0.1724	0.2848	堅持	18	0.1865	0.3120	制	19	0.2634	0.4096	人類	13	0.1408	0.2412
48	族	45	0.1724	0.2848	問題	17	0.1761	0.2946	交流	19	0.2634	0.4096	時代	13	0.1408	0.2412
49	文明	43	0.1647	0.2722	成為	17	0.1761	0.2946	和諧	19	0.2634	0.4096	更	13	0.1408	0.2412
50	中華	43	0.1647	0.2722	現代化	17	0.1761	0.2946	推動	19	0.2634	0.4096	人類	13	0.1408	0.2412
51	岸	43	0.1647	0.2722	進一步	17	0.1761	0.2946	關係	18	0.2495	0.3880	歷史	13	0.1408	0.2412
52	地區	42	0.1609	0.2659	美好	17	0.1761	0.2946	統一	18	0.2495	0.3880	即將	12	0.1299	0.2227
53	保持	42	0.1609	0.2659	不斷	17	0.1761	0.2946	生活	18	0.2495	0.3880	工作	12	0.1299	0.2227
54	世紀	42	0.1609	0.2659	事業	17	0.1761	0.2946	增長	17	0.2357	0.3665	澳門	11	0.1191	0.2041
55	符村	41	0.1571	0.2595	葡門	16	0.1658	0.2773	里	17	0.2357	0.3665	地區	11	0.1191	0.2041
56	推動	41	0.1571	0.2595	積極	16	0.1658	0.2773	臺灣	16	0.2218	0.3449	致以	11	0.1191	0.2041
57	加融	40	0.1532	0.2532	地區	16	0.1658	0.2773	朋友們	15	0.2079	0.3233	不少	11	0.1191	0.2041
58	葡門	40	0.1532	0.2532	保持	16	0.1658	0.2773	地區	15	0.2079	0.3233	困難	11	0.1191	0.2041
59	歷史	40	0.1532	0.2532	次	16	0.1658	0.2773	挑戰	14	0.1941	0.3018	經濟	11	0.1191	0.2041
60	事業	40	0.1532	0.2532	生活	16	0.1658	0.2773	事業	14	0.1941	0.3018	推動	11	0.1191	0.2041
61	維護	40	0.1532	0.2532	廣播	15	0.1554	0.2600	實現	14	0.1941	0.3018	生命	11	0.1191	0.2041
62	偉大	39	0.1494	0.2469	電臺	15	0.1554	0.2600	幸福	14	0.1941	0.3018	小康	11	0.1191	0.2041
63	中央	38	0.1456	0.2405	特色	15	0.1554	0.2600	變化	14	0.1941	0.3018	農村	11	0.1191	0.2041
64	更加	38	0.1456	0.2405	前進	15	0.1554	0.2600	政策	13	0.1802	0.2802	脫貧	11	0.1191	0.2041
65	取得	37	0.1418	0.2342	方針	15	0.1554	0.2600	著力	13	0.1802	0.2802	首	11	0.1191	0.2041
66	方針	37	0.1418	0.2342	關係	15	0.1554	0.2600	全	12	0.1663	0.2587	習近平	10	0.1083	0.1856
67	政治	36	0.1379	0.2279	岸	15	0.1554	0.2600	項	12	0.1663	0.2587	朋友們	10	0.1083	0.1856
68	希望	36	0.1379	0.2279	創造	15	0.1554	0.2600	小康	12	0.1663	0.2587	行政區	10	0.1083	0.1856
69	廣播	35	0.1341	0.2215	中央	15	0.1554	0.2600	進程	12	0.1663	0.2587	積極	10	0.1083	0.1856
70	電臺	35	0.1341	0.2215	海外	14	0.1450	0.2426	願	12	0.1663	0.2587	辭解人	10	0.1083	0.1856
71	關係	34	0.1303	0.2152	更	14	0.1450	0.2426	最後	12	0.1663	0.2587	建成	10	0.1083	0.1856
72	創造	34	0.1303	0.2152	未來	14	0.1450	0.2426	科學	12	0.1663	0.2587	目標	10	0.1083	0.1856
73	北京	33	0.1264	0.2089	繁榮	14	0.1450	0.2426	快	11	0.1525	0.2371	疫情	10	0.1083	0.1856
74	進步	33	0.1264	0.2089	偉大	14	0.1450	0.2426	中	11	0.1525	0.2371	很多	10	0.1083	0.1856
75	重要	33	0.1264	0.2089	共產黨	14	0.1450	0.2426	作出	11	0.1525	0.2371	廣	10	0.1083	0.1856
76	制	33	0.1264	0.2089	團結	13	0.1347	0.2253	掛職	11	0.1525	0.2371	首	10	0.1083	0.1856
77	致以	32	0.1226	0.2026	項	13	0.1347	0.2253	加快	11	0.1525	0.2371	繁榮	10	0.1083	0.1856
78	海外	31	0.1188	0.1962	完成	13	0.1347	0.2253	北京	11	0.1525	0.2371	幸福	9	0.0975	0.1670
79	不懈	31	0.1188	0.1962	希望	13	0.1347	0.2253	條條不離	11	0.1525	0.2371	幹部	9	0.0975	0.1670
80	幸福	30	0.1149	0.1899	族	13	0.1347	0.2253	胡錦濤	10	0.1386	0.2156	取得	9	0.0975	0.1670
81	黨	29	0.1111	0.1836	金融	13	0.1347	0.2253	鐘聲	10	0.1386	0.2156	深化	9	0.0975	0.1670
82	更	29	0.1111	0.1836	更加	13	0.1347	0.2253	即將	10	0.1386	0.2156	事業	9	0.0975	0.1670
83	項	29	0.1111	0.1836	推進	13	0.1347	0.2253	高興	10	0.1386	0.2156	次	9	0.0975	0.1670
84	共產黨	29	0.1111	0.1836	高興	12	0.1243	0.2080	電視臺	10	0.1386	0.2156	勝利	9	0.0975	0.1670
85	交流	29	0.1111	0.1836	建立	12	0.1243	0.2080	海外	10	0.1386	0.2156	進行	9	0.0975	0.1670
86	現代化	28	0.1073	0.1772	政府	12	0.1243	0.2080	僑胞	10	0.1386	0.2156	屆	9	0.0975	0.1670
87	挑戰	28	0.1073	0.1772	政策	12	0.1243	0.2080	致以	10	0.1386	0.2156	心	9	0.0975	0.1670
88	進一步	28	0.1073	0.1772	外交	12	0.1243	0.2080	代表	10	0.1386	0.2156	慶	9	0.0975	0.1670
89	次	28	0.1073	0.1772	擴大	12	0.1243	0.2080	重要	10	0.1386	0.2156	家	9	0.0975	0.1670
90	小康	28	0.1073	0.1772	致以	11	0.1140	0.1906	災害	10	0.1386	0.2156	徵	9	0.0975	0.1670
91	會	27	0.1034	0.1709	困難	11	0.1140	0.1906	文明	10	0.1386	0.2156	2020年	9	0.0975	0.1670

순위	ALL	F	T(%)	C(%)	JIANG	F	T(%)	C(%)	HU	F	T(%)	C(%)	XI	F	T(%)	C(%)
92	週年	27	0.1034	0.1709	貢獻	11	0.1140	0.1906	高	10	0.1386	0.2156	事	9	0.0975	0.1670
93	時代	27	0.1034	0.1709	獨立自主	11	0.1140	0.1906	舉	10	0.1386	0.2156	同志們	8	0.0866	0.1485
94	增長	27	0.1034	0.1709	環境	11	0.1140	0.1906	進一步	10	0.1386	0.2156	族	8	0.0866	0.1485
95	即將	27	0.1034	0.1709	秩序	11	0.1140	0.1906	國家	10	0.1386	0.2156	成就	8	0.0866	0.1485
96	政策	26	0.0996	0.1646	制	11	0.1140	0.1906	友好	10	0.1386	0.2156	謝謝	8	0.0866	0.1485
97	治	26	0.0996	0.1646	全面	11	0.1140	0.1906	發掘	10	0.1386	0.2156	重大	8	0.0866	0.1485
98	特色	25	0.0958	0.1582	國民	11	0.1140	0.1906	港人	10	0.1386	0.2156	奉獻	8	0.0866	0.1485
99	困難	25	0.0958	0.1582	精神	11	0.1140	0.1906	港	10	0.1386	0.2156	合作	8	0.0866	0.1485
100	未來	25	0.0958	0.1582	推動	11	0.1140	0.1906	澳人	10	0.1386	0.2156	活動	8	0.0866	0.1485

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑, ‘F’: 빈도, ‘T’: 총어휘수대비비율, ‘C’: 총내용어수대비비율

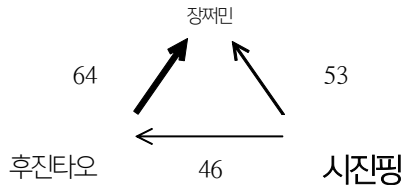
〈표 4〉 좌측에 격자로 채워진 부분이 지도자 통합 상위 100개 목록이고 그중에서 배경색이 들어간 부분은 세 지도자가 공히 사용한 어휘를 가리킨다. 지도자별 개별 어휘목록을 보면, 총 내용어수 대비 1% 이상을 점하는 어휘가 후진타오 11개(1위~11위), 장쩌민 9개(1위~9위), 시진핑 4개(1위~4위) 순이고 1.0% 이하, 0.5% 이상을 점하는 어휘가 후진타오 21개(12위~32위), 장쩌민은 17개(10위~26위), 시진핑은 5개(5위~9위) 순이다. 이를 통합하면, 지도자별 100개 어휘 중에서 내용어 대비 0.5% 이상을 차지하는 어휘수는 후진타오 32개, 장쩌민 26개, 시진핑 9개 순으로 시진핑 이후로 고빈도 어휘 구간이 급격히 좁아지고 있는 것을 볼 수 있다. 이는 시진핑의 신년사가 기타 두 지도자와 비교해 일부 소수의 고빈도 어휘군에 의존하기보다는 (상대적으로) 다양하게 분포된 저빈도 어휘들로 구성됐을 것이란 가정을 하게 한다. 실제로, 상위 1% 이상의 어휘들이 전체 내용어에서 점한 비율의 합은 후진타오 21.1036%, 장쩌민 18.1803%, 시진핑 6.0865%이다. 범위를 0.5% 이상으로 넓히면 후진타오 35.7621%, 장쩌민 29.0123%, 시진핑 10.0389% 순으로 [후진타오, 장쩌민]과 [시진핑] 간의 격차는 더 커진다. 이것은 앞서 TTR과 평균문장수로부터 관찰된, 시진핑의 신년사는 어휘다양도와 정보량 면에서 장쩌민, 후진타오보다 높다는 해석과 같은 연장선에 있다. 마지막으로, 지도자 간 어휘 교차량을 살펴본다.

〈표 5〉 지도자 간 상위 100개 내 어휘 일치수

지도자	JIANG	HU	XI
JIANG	100	64	53
HU	64	100	46
XI	53	46	100

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑

교차사용 면에서 〈표5〉는 장쩌민과 후진타오 간에 64개 어휘가 일치하고, 후진타오와 시진핑 간에 46개, 시진핑과 장쩌민 간에 53개 어휘가 일치한다는 것을 보여준다. 이는 〈그림 1〉과 같이 지도자 3인의 어휘 교차 양상으로 표현 가능하다.



〈그림 1〉 지도자간 신년사 어휘 교차

〈그림 1〉은 선분의 굵기, 화살의 방향, 지도자명의 크기로 어휘교차 상황을 표시하고 있다. 이 중에서 선분의 굵기는 어휘간 교차량을 표시한다. 장쩌민의 어휘는 후진타오와 64개, 시진핑과 53개 교차하며 11개의 차를 보인다. 후진타오의 어휘는 시진핑과 46개, 장쩌민과 64개 교차하며 18개 차를 보인다. 그리고 시진핑의 어휘는 장쩌민과 53개, 후진타오와 46개 교차하며 7개 차를 보인다. 화살표는 어휘교차의 실행 방향을 나타낸다. 장쩌민과 후진타오 간 어휘교차량은 후진타오에 의해 좌우된다. 장쩌민은 퇴임자이므로 신년사 어휘의 교차 권한이 주어지지 않기 때문이다. 마찬가지로 시진핑을 중심으로 한 후진타오와 장쩌민 간 교차량도 시진핑에 의해 서만 좌우된다. 그리고 이를 통해 발생한 권한의 중첩은 시진핑, 후진타오,

장쩌민 순의 지도자명 크기로 표시된다.

이상의 어휘교차량, 교차의 실행방향, 권한중첩의 정도를 종합해 보면 <표5>의 가장 큰 기여자는 시진핑이라고 할 수 있다. 그리고 장쩌민과 53개, 후진타오와 46개 일치어를 보이며 (가장 적은)7개 차를 보인 시진핑의 어휘교차 결과는 전임 지도자들과 거리를 두고자 한 시진핑의 언어사용 전략으로 해석할 수 있다.

IV. 신년사 핵심어와 시대 키워드

1. 신년사의 핵심어 비교

흔히 출현빈도가 높을수록 중요한 어휘, 즉 핵심어로 여겨진다.¹⁴⁾ 중요하기 때문에 많이 사용되었을 것이란 가정도 일리 있으나 모든 화자가 맥락 불문하고 많은 빈도로 사용했다면 그 어휘가 정말 중요해서 사용한 것인지 아니면 관례적으로 사용해야 했던 것인지 판단하기 쉽지 않다. TF-IDF는 여기에 참고점을 제공하는 통계적 방법이다.

TF-IDF는 ‘Term Frequency-Inverse Document Frequency’의 준말로 흔히 ‘어휘/용어/단어빈도-역문서 빈도’로 불린다. 이 중 TF는 어휘 A가 문서 전체에 걸쳐 출현한 빈도를, IDF는 그 어휘가 출현한 문서수에 반비례하는 가중치를 가리킨다. TF-IDF는 기본적으로, TF에 IDF를 곱함으로써 출현한 문서수가 많아질수록 어휘의 전체빈도를 상쇄시키고 특정 또는 일부 문서에 출현했다면 그것의 희소성을 인정하여 가중치를 부여하는 방식이다. 이를 통해, 전체 지도자들이 중요시한 어휘는 물론, 개별 지도자들이 특별한 가치를 부여한 핵심어를 가늠할 수 있다. TF-IDF의 수식은 다음

14) 본고에서는 지도자별로 중요시한 어휘를 ‘(신년사)핵심어’로, 시대별로 중요시된 어구를 ‘(시대)키워드’로 지칭한다.

과 같다.

$$(3) TF-IDF(t, d) = TF(t, d) \times IDF(t)$$

$$\cdot \text{어휘빈도: } TF(t, d) = \frac{\text{문서 } d \text{ 내에서 단어 } t \text{ 가 출현한 횟수}}{\text{문서 } d \text{ 의 총 내용어수}}$$

$$\cdot \text{역문서빈도: } IDF(t) = \log \frac{N}{DF_t} + 1$$

장쩌민의 中國을 예로 들어보자. TF(t, d)에서 t는 단어로써 中國, d는 문서로서 장쩌민 신년사를 가리킨다. 그러므로 어휘빈도 TF(t, d)는 ‘장쩌민 신년사에서 中國이 출현한 횟수’를 ‘장쩌민 신년사의 총내용어수’로 나눈 값이다. 출현횟수(절대빈도) 자체가 아닌, 출현횟수를 총내용어수로 나눔으로써 코퍼스 크기에 따른 출현횟수비율(상대빈도)로 TF(t, d)를 산출한다. 다음으로, 역문서빈도 IDF(t)에서 분자 N은 총문서수, 분모 DF_t는 t(中國)가 출현하는 문서수를 가리킨다. 본고에서는 장쩌민, 후진타오, 시진핑의 신년사 각각을 하나의 ‘문서’로 간주하고 N에 3을 부여한다. 그리고 DF_t에도 中國이 장쩌민, 후진타오, 시진핑 신년사에 모두 출현하므로 3을 놓는다. 이에 TF(t, d)와 IDF(t)의 곱은 다음과 같다.

$$(4) TF-IDF(t, d) = TF\left(\frac{228}{5807}\right) \times IDF\left[\log\left(\frac{3}{3}\right)+1\right]$$

$$\therefore \text{장쩌민의 ‘中國’ } TF-IDF \approx 0.039263$$

역문서빈도(IDF_t)값이 1이면 핵심어가 모든 문서에 출현함을 가리킨다. 반면에 1보다 크면 클수록 ‘지도자 특정한’인 어휘로 간주한다. IDF가 1보다 크다는 것은 해당 어휘가 오직 또는 일부 신년사에서만 사용됐을 수 있음을 의미하기 때문이다.

아래는 TF-IDF를 통해 산출된 장쩌민, 후진타오, 시진핑의 핵심어 상위 50개이다.

〈표 6〉 중국지도자별 신년사 핵심어 상위 50개

순위	JIANG	빈도	TF	IDF	TF-IDF	HU	빈도	TF	IDF	TF-IDF	XI	빈도	TF	IDF	TF-IDF
1	中國	228	0.0395	1.0000	0.0395	發財	141	0.0304	1.0000	0.0304	中國	105	0.0195	1.0000	0.0195
2	人民	143	0.0248	1.0000	0.0248	中國	126	0.0272	1.0000	0.0272	人民	87	0.0161	1.0000	0.0161
3	發展	136	0.0236	1.0000	0.0236	人民	120	0.0259	1.0000	0.0259	年	69	0.0128	1.0000	0.0128
4	世界	116	0.0201	1.0000	0.0201	和平	103	0.0222	1.0000	0.0222	國	67	0.0124	1.0000	0.0124
5	國	112	0.0194	1.0000	0.0194	世界	99	0.0213	1.0000	0.0213	世界	48	0.0089	1.0000	0.0089
6	和平	86	0.0149	1.0000	0.0149	國	99	0.0213	1.0000	0.0213	新	47	0.0087	1.0000	0.0087
7	新	81	0.0140	1.0000	0.0140	經濟	77	0.0166	1.0000	0.0166	發展	45	0.0084	1.0000	0.0084
8	年	79	0.0137	1.0000	0.0137	國際	62	0.0134	1.0000	0.0134	改革	39	0.0072	1.0000	0.0072
9	經濟	68	0.0118	1.0000	0.0118	新	55	0.0119	1.0000	0.0119	全面	34	0.0063	1.0000	0.0063
10	世紀	41	0.0071	1.4055	0.0100	合作	50	0.0108	1.0000	0.0108	同胞	26	0.0048	1.0000	0.0048
11	國際	54	0.0094	1.0000	0.0094	共同	47	0.0101	1.0000	0.0101	國家	25	0.0046	1.0000	0.0046
12	建設	45	0.0078	1.0000	0.0078	同胞	45	0.0097	1.0000	0.0097	全	25	0.0046	1.0000	0.0046
13	繼續	44	0.0076	1.0000	0.0076	繼續	45	0.0097	1.0000	0.0097	貧困	25	0.0046	1.0000	0.0046
14	共同	39	0.0068	1.0000	0.0068	兩	44	0.0095	1.0000	0.0095	實現	25	0.0046	1.0000	0.0046
15	兩	39	0.0068	1.0000	0.0068	社會	43	0.0093	1.0000	0.0093	新年	24	0.0045	1.0000	0.0045
16	中	38	0.0066	1.0000	0.0066	年	40	0.0086	1.0000	0.0086	脫貧	11	0.0020	2.0986	0.0043
17	社會主義	36	0.0062	1.0000	0.0062	建設	37	0.0080	1.0000	0.0080	農村	11	0.0020	2.0986	0.0043
18	人類	35	0.0061	1.0000	0.0061	促進	37	0.0080	1.0000	0.0080	生命	11	0.0020	2.0986	0.0043
19	社會	35	0.0061	1.0000	0.0061	推進	34	0.0073	1.0000	0.0073	共同	23	0.0043	1.0000	0.0043
20	國家	34	0.0059	1.0000	0.0059	香港	34	0.0073	1.0000	0.0073	群眾	23	0.0043	1.0000	0.0043
21	改革	33	0.0057	1.0000	0.0057	堅持	34	0.0073	1.0000	0.0073	生活	23	0.0043	1.0000	0.0043
22	民族	32	0.0055	1.0000	0.0055	澳門	32	0.0069	1.0000	0.0069	社會	22	0.0041	1.0000	0.0041
23	臺灣	31	0.0054	1.0000	0.0054	方針	22	0.0047	1.4055	0.0067	號	15	0.0028	1.4055	0.0039
24	同胞	30	0.0052	1.0000	0.0052	電臺	20	0.0043	1.4055	0.0061	很多	10	0.0019	2.0986	0.0039
25	文明	30	0.0052	1.0000	0.0052	廣播	20	0.0043	1.4055	0.0061	努力	20	0.0037	1.0000	0.0037
26	香港	29	0.0050	1.0000	0.0050	開放	28	0.0060	1.0000	0.0060	人口	14	0.0026	1.4055	0.0037
27	實現	28	0.0049	1.0000	0.0049	特別	28	0.0060	1.0000	0.0060	美好	19	0.0035	1.0000	0.0035
28	促進	28	0.0049	1.0000	0.0049	行政區	28	0.0060	1.0000	0.0060	奮鬥	19	0.0035	1.0000	0.0035
29	合作	28	0.0049	1.0000	0.0049	和諧	19	0.0041	1.4055	0.0058	黨	19	0.0035	1.0000	0.0035
30	統一	26	0.0045	1.0000	0.0045	全面	25	0.0054	1.0000	0.0054	繼續	19	0.0035	1.0000	0.0035
31	穩定	25	0.0043	1.0000	0.0043	社會主義	25	0.0054	1.0000	0.0054	2020年	9	0.0017	2.0986	0.0035
32	里	25	0.0043	1.0000	0.0043	努力	25	0.0054	1.0000	0.0054	心	9	0.0017	2.0986	0.0035
33	全	25	0.0043	1.0000	0.0043	族	24	0.0052	1.0000	0.0052	幹部	9	0.0017	2.0986	0.0035
34	祖國	24	0.0042	1.0000	0.0042	穩定	24	0.0052	1.0000	0.0052	家	9	0.0017	2.0986	0.0035
35	政治	24	0.0042	1.0000	0.0042	維護	24	0.0052	1.0000	0.0052	偉大	18	0.0033	1.0000	0.0033
36	進步	24	0.0042	1.0000	0.0042	問題	24	0.0052	1.0000	0.0052	香港	18	0.0033	1.0000	0.0033
37	歷史	24	0.0042	1.0000	0.0042	改革	22	0.0047	1.0000	0.0047	國際	17	0.0032	1.0000	0.0032
38	朋友們	23	0.0040	1.0000	0.0040	保持	22	0.0047	1.0000	0.0047	推進	17	0.0032	1.0000	0.0032
39	中華	22	0.0038	1.0000	0.0038	中央	22	0.0047	1.0000	0.0047	民族	17	0.0032	1.0000	0.0032
40	開放	22	0.0038	1.0000	0.0038	新年	21	0.0045	1.0000	0.0045	和平	17	0.0032	1.0000	0.0032
41	方針	15	0.0026	1.4055	0.0037	岸	21	0.0045	1.0000	0.0045	奉獻	8	0.0015	2.0986	0.0031
42	廣播	15	0.0026	1.4055	0.0037	繁榮	21	0.0045	1.0000	0.0045	負	8	0.0015	2.0986	0.0031
43	電臺	15	0.0026	1.4055	0.0037	澳人	10	0.0022	2.0986	0.0045	中	16	0.0030	1.0000	0.0030
44	努力	21	0.0036	1.0000	0.0036	發財觀	10	0.0022	2.0986	0.0045	祖國	16	0.0030	1.0000	0.0030
45	取得	21	0.0036	1.0000	0.0036	治	20	0.0043	1.0000	0.0043	祝福	16	0.0030	1.0000	0.0030
46	新年	21	0.0036	1.0000	0.0036	人類	20	0.0043	1.0000	0.0043	週年	16	0.0030	1.0000	0.0030
47	1995年	10	0.0017	2.0986	0.0036	加勝	20	0.0043	1.0000	0.0043	百	11	0.0020	1.4055	0.0029
48	加勝	19	0.0033	1.0000	0.0033	積極	19	0.0041	1.0000	0.0041	更加	15	0.0028	1.0000	0.0028
49	澳門	19	0.0033	1.0000	0.0033	推動	19	0.0041	1.0000	0.0041	建設	15	0.0028	1.0000	0.0028
50	重要	19	0.0033	1.0000	0.0033	交流	19	0.0041	1.0000	0.0041	開放	15	0.0028	1.0000	0.0028

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑

〈표 6〉의 각 50개 어휘는 장쩌민, 후진타오, 시진핑이 각 10년 동안 신

년사에서 빈번하게, 그리고 강조하여 사용한 핵심어라고 할 수 있다. 먼저, TF의 관점에서 TF-IDF값을 살펴보자. 상위 10위 중에서 ‘中國(J: 0.0395, H: 0.0272, X: 0.0195)’, ‘人民(J: 0.0248, H: 0.0259, X: 0.0161)’, ‘發展(J: 0.0236, H: 0.0304, X: 0.0084)’, ‘世界(J: 0.0201, H: 0.0213, X: 0.0089)’, ‘國(J: 0.0194, H: 0.0213, X: 0.0124)’, ‘新(J: 0.0140, H: 0.0119, X: 0.0087)’은 장쩌민, 후진타오, 시진핑에 이르는 30년 신년사에서 꾸준히 사용돼 온 핵심어라고 할 수 있다. 특히, ‘國’와 ‘新’를 제외한 ‘中國’, ‘人民’, ‘發展’, ‘世界’는 중국지도자 신년사의 공통 핵심어로서 지위가 굳건하다. 다만, 이들 핵심어의 TF-IDF가 [JIANG, HU] 對 [XI]의 구도 속에서 큰 폭으로 하락하고 있는 점은, 핵심어들의 지위는 유지되고 있으나 중요도는 신년사라는 맥락 속에서 약화 추세에 있음을 보여준다.

다음으로 IDFt의 관점에서 살펴보자. DFt가 1.4055라면(〈표 6〉의 열은 회색) 해당 어휘가 2명의 연설에서, 2.0986이라면(〈표 6〉의 짙은 회색) 1명의 연설에서 사용됐을 가능성이 높다. 이를 참고하면, 핵심어 50개 중에서 장쩌민만 사용했을 가능성이 높은 어휘는 ‘1995年’ 뿐이다. 나머지 ‘世紀’, ‘方針’, ‘廣播’, ‘電臺’는 후진타오 또는 시진핑과 함께 사용한 핵심어이다. 후진타오만의 핵심어로는 ‘澳人’과 ‘發展觀’이 있고 나머지 ‘方針’, ‘電臺’, ‘廣播’, ‘和諧’는 시진핑 또는 장쩌민과 함께 사용한 핵심어이다. 반면에, 시진핑 신년사에서는 ‘號’, ‘人口’, ‘百’을 제외한, ‘脫貧’, ‘農村’, ‘生命’, ‘很多’, ‘2020年’, ‘心’, ‘幹部’, ‘家’, ‘奉獻’, ‘負’가 단독으로 사용된 핵심어로서 과거 지도자들에 비해 늘어나 있다.

장쩌민에게 특히 중요시된 ‘1995年’은 항일전쟁 전승 50주년, UN성립 50주년 기념 연설, 제8차 경제5개년계획(마지막 해), 덩샤오핑(鄧小平)이 목표했던 세기말 GDP 달성 등과 관련된 해로서 국제사회 속 국가 위상 제고와 경제 개혁/발전의 지속성이 강조된 해였음을 보여준다. ‘廣播’, ‘電臺’는 장쩌민과 후진타오 신년사에만 출현하는 어휘이다. 두 지도자가 ‘通過中國國際廣播電臺、中央人民廣播電臺和中央電視臺’를 신년사 서두에서 20년간 의례적으로 사용해왔기 때문이다. 이를 친매체적이라고 볼 수도

있겠으나 시진핑의 신년사 이후로 언급되지 않은 것은 오히려 환경 변화에 기인한 바가 크다고 본다. 스마트폰에 의한 인터넷 사용 및 영상 시청의 증가로 인해 더 이상 텔레비전과 라디오만을 담화 발표의 유일 경로로 볼 수 없기 때문이다. 이에 ‘廣播’, ‘電臺’는 장쩌민과 후진타오의 핵심어인 동시에 담화 매체 환경에 커다란 변화가 있었음을 보여주는 시대 반영어이기도 하다. 후진타오의 핵심어인 ‘發展觀’, 그리고 시진핑과 함께 사용한 ‘和諧’는 후진타오의 사상 및 정책기조를 보여주는 핵심어이다. 후진타오는 경제발전 과정에서 발생한 양극화 해소를 위해 재임기간 동안 사람 중심의 균형잡힌 발전과 지속가능한 발전을 도모하며(科學發展觀)¹⁵⁾ 조화로운 사회(和諧社會)¹⁶⁾를 건설하고자 했는데, 신년사 속 ‘發展觀’과 ‘和諧’는 이에 대한 반영으로 볼 수 있다.

시진핑의 ‘2020年’은 3년(2017년, 2019년, 2020년)에 걸쳐 출현하였다. ‘2020年’하면 흔히 COVID19을 떠올리기 쉽지만 시진핑 신년사에서 ‘2020年’은 ‘脫貧’, ‘農村’과 맥락을 같이한다. ‘2020年’까지 ‘脫貧’하겠다고 2017년 신년사에서 약속한 이후 2019년, 2020년 연이어 그 약속과 목표 달성을 재확인하기 때문이다. 또한 ‘幹部’는 시진핑 신년사에 넓게 분포된 어휘로서 2013년 임기 첫 해부터 사용된 핵심어이다. 시진핑의 ‘幹部’는 반부패와 기율 등 국가 기강과 관련될 것 같지만, 신년사에서는 오히려 국민의 ‘生命’을 위해 그리고 국정 목표인 ‘脫貧’과 소강사회 건설을 위해 ‘奉獻’하는 당의 리더들로 지칭된다. 그로 인해 ‘幹部’는 ‘扶貧’, ‘基層’, ‘駐村’ 등의 수식어와 자주 호응을 이루며 보통 치하(致賀)의 맥락 속에서 등장한다.

정리하면, 신년사라는 맥락 속에서 장쩌민과 후진타오는 자신만의 핵심어보다 기존의 핵심어를 사용하는 관례적 경향이 큰 반면 시진핑은 기존 공통 핵심어와 함께 자신만의 개인적인 핵심어도 더불어 부각시키는 경향

15) Wikiwand(科學發展觀): <https://www.wikiwand.com/zh/articles/科學發展觀>.

16) Wikiwand(和諧社會): <https://www.wikiwand.com/zh/articles/和諧社會>.

을 띠고 있다.

2. 신년사와 시대 키워드

앞서서는 신년사라는 맥락 속에서 장쩌민, 후진타오, 시진핑 어휘의 TF-IDF를 살펴보고 TF와 IDF값에 기초하여 지도자 간 또는 지도자별 핵심어가 갖는 의미와 추세에 대해 논했다.

대상을 전환하여, 신년사와 시대기조 간의 연관성에 대해 살펴본다. 국가별로 정부가 풀어내야 하는 시대적 과제들이 존재한다. 그리고 그 과제들을 이행 또는 홍보하는 과정에서 시대별로 회자되는 이념명, 정책명, 구호 등이 더불어 생산된다. 흔히 ‘(시대)키워드’로 통칭되는 이들 어휘(어구)는 신년사 속에서 있는 그대로 사용될 때도 있지만 키워드에 관련된 배경 설명이나 정책 관련 언급 등을 통해 신년사와 간접 연관성을 맺기도 한다. 그 결과 시대 키워드가 신년사에 언급되지 않거나(예: ‘醫療改革’) 언급됐어도 분절(segmentation)과정에서 하나의 단위로 인식되지 못한 경우(예: ‘三個代表重要思想’)가 발생한다.

이에 대해, 본고는 문자 단위로 키워드나 신생어를 처리하는 문자 엔그램(character n-grams)방법을 적용한다. 단어 구분이 공백에 의존하지도 않고 단어와 형태소 모두가 단음절로 존재할 수 있는 중국어는 자연어 분절에서 어려움을 갖는다.¹⁷⁾ 문자 엔그램은 단위를 문자에 뭉으로써 중국어의 이 같은 난점을 극복하는데 토대를 제공한다.¹⁸⁾ 본고는 단어분절 라이브러리로 HanLP를 사용했는데 HanLP의 분절 결과가 시대 키워드와 완전하게 일치하지 못할 경우, 글자 한 자 단위의 유니그램(unigram)을 이용하여 키워드와 신년사 간의 연관성 관찰을 시도했다. 그 과정은 다음과 같다.

17) Mikheev, Andrei, “Text Segmentation”, in Mitkov, Ruslan ed., *The Oxford Handbook of Computational Linguistics* (2nd edition), Oxford University Press, 2022, 557쪽.

18) 상동, 558쪽.

먼저 30년 간의 시대 키워드를 연도별로 산출한다. 그리고 그것을 10년 단위로 묶어 중복 키워드는 배제하고 장쩌민, 후진타오, 시진핑 기간만의 키워드 목록으로 만든다. 키워드가 신년사 내의 내용어와 완전히 일치하면 해당 어휘의 TF-IDF를 그대로 반영한다. 하지만 키워드에 상응하는 어휘가 없다면 신년사 속 개별글자로 키워드를 구성하고 각각의 개별글자가 가진 TF-IDF를 합산한다. 키워드의 TF-IDF가 신년사의 문자 엔그램 결과라는 점에서 TF-IDF 합산 점수가 높은 키워드일수록 신년사에서 중시되고 신년사와의 연관성도 높을 것이라 판단한다.

30년간의 연도별 ‘시대 키워드’는 Perplexity¹⁹⁾를 이용하여 추출하였다. 프롬프트의 예는 다음과 같다.

〈표 7〉 연도별 시대 키워드 추출 프롬프트

```
@학술 .....
>> 아래 내용과 관련한 논문과 서적, 매체 등을 찾아서 핵심 연관 키워드를 **연도별로 20개씩**
아래 <format>으로 정리해줘.
1. 江澤民(1993-2002), 胡錦濤(2003-2012), 習近平(2013-2022) 각각의 집권 동안, 각 시기별로 강
조한 중요 정치사상, 이념, 정책, 구호 등
2. 江澤民(1993-2002), 胡錦濤(2003-2012), 習近平(2013-2022) 각각의 집권 동안 발생했던 각종
사건, 사고, 국내 행사, 국제 행사, 제양 등
3. 江澤民(1993-2002), 胡錦濤(2003-2012), 習近平(2013-2022) 집권 동안 이 세 지도자가 자국의
국민들에게 강조했던 또는 강하게 전달하려 했던 메시지
4. 江澤民(1993-2002), 胡錦濤(2003-2012), 習近平(2013-2022) 집권 동안 이 세 지도자가 香港, 澳
門, 臺灣 국민들에게 강조했던 또는 강하게 전달하려 했던 메시지
<response format>
1. 江澤民(1993-2002):
1.1 江澤民(1993) : [연관 키워드1, 연관 키워드2, 연관 키워드3, ..., 연관 키워드20]
.....
</response format>
```

이로써 얻은 연도별 시대 키워드(연도별 20개씩하여 지도자별 200개, 총합 600개)는 각각 중복 제거, 지도자별 통합을 통해 다음과 같이 재임기

19) Perplexity: <https://www.perplexity.ai/>

간별 시대 키워드로 정리된다.

(5) - 江澤民(1993-2002, 총 117개):

民主建設, 經濟調整, 教育改革, 環保問題, 臺灣問題, 經濟全球化, 農村改革, 醫療改革, 經濟增長, 城鄉發展, 兩個一百年, 三個代表, 法制深化, 國際影響力, 法治中國, 民族政策, 發展開放區, 文化發展, 神舟飛船, 產業升級, 教育發展, 環保建設, 經濟結構調整, 社會保障, 社會穩定, 農村振興, 對外開放, 教育強國, 科學發展, 對臺政策, WTO支持, 法輪功, 經濟快速增長, 民族團結, 對外經濟合作, 臺灣統一, WTO加入, 民生改善, 教育現代化, 文化繁榮, 社會主義市場經濟, 現代化建設, 貿易開放, 法制建設, 香港回歸, 國防科技, 法制完善, 國企重組, 環境保護, 環保升級, 臺灣關係, 中美關係, 亞洲金融危機, 科學技術發展, 民主法治, 經濟快速發展, 國防現代化, 發展農村, 科技創新, 法制進步, 海外華人, 港澳事務穩定, 農村經濟, 地方經濟, 臺灣和平談判, 腐敗整治, 香港政策, 農村發展, 城鄉差距縮小, 政治體制改革, 醫療保障, 生態環保, WTO加入準備, 城鄉統籌, 東西部開發, 北京申奧, WTO準備, 科技進步, 中國加入WTO, 科教興國, WTO談判, 臺海局勢, 科技興國, 國企改革, 社會保障完善, 臺灣政策, 港澳穩定, 國際貿易, 朱鎔基, 經濟發展, 軍隊經商取消, 跨國投資, 現代國防, 社會改革, 國防升級, 科學技術, 文化建設, 經濟穩定增長, 腐敗治理, 科學發展觀鋪墊, 985工程, 改革開放, 社會保障強化, 科技研發, 國際合作, 高校建設, 環保政策, 軍隊現代化, 高新技術發展, 綠色環保, 港澳事務, 軍隊改革, 澳門政策, 國際關係穩定, 澳門回歸, 臺灣統一論, 社會保障改革

- 胡錦濤(2003-2012, 총 89개):

經濟轉型, 經濟調整, 北京奧運, 教育改革, 反腐敗鬥爭, 臺灣問題, 中央軍委, 政治改革, 醫療改革, 醫療體系, 經濟增長, 和諧社會, 城鄉平衡, 上海世博, 臺灣海峽, 港澳問題, 法治政府, 文化發展, 農業振興, 住房政策, 教育發展, 反分裂國家法, 經濟結構調整, 社會保障, 社會穩定, 農村振興, 資源節約, 法治建設, 國際交流, 法治進展, 文化普及, 國際關係, 依法治國, 法治進步, 教育振興, 臺灣統一, 反恐鬥爭, 教育現代化, 文化繁榮, 法治完善, 港澳政策, 反恐怖, 醫療體制, 反腐加強, 法制建設, 文化提升, 反腐敗, 法治推進, 環境保護, 醫療改善, 中美關係, 汶川地震, 住房改革, 經濟復蘇, 軍隊建設, 科學發展觀, 生態保護, 反腐鬥爭, 反腐倡廉, 衛生改革, 農村經濟, 農業發展, 農村發展, 醫療保障, 生態文明, 反恐安全, 住房保障, 醫療衛生, 城鄉差距, 資源保護, 臺灣政策, 港澳穩定, 環境政策, 經濟發展, 文化建設, 國際合作, 社會主義和諧社會, 環保政策, 軍隊現代化, 城鄉協調, 人民生活, 生態環境, 法制推進, 反恐體系, 臺灣和平發展, 港澳事務, 軍隊改革, 和諧社會構建, SARS

- 習近平(2013-2022, 총 49개):

全面深化改革, 經濟轉型, 經濟復蘇, 教育改革, 反腐敗鬥爭, 臺灣問題, 依法治國, 二十大, 新冠疫情, 憲法修改, 反腐倡廉, 中國夢, 國家安全, 科技創新, 文化自信, 新能源, 國際影響力, 綠色發展, 科技自立, 人民為中心, 澳門穩定, 內外開放, 一帶一路, 臺灣統一, 澳門一國兩制, 國際合作, 脫貧攻堅, 十九大, 香港治理, 高質量發展, 生態文明, 創新驅動, 科技強國, 軍隊現代化, 共同富裕, 香港一國兩制, 反腐敗, 新常態經濟, 香港事務, 民族復興, 臺灣政策, 澳門事務, 新冠疫情防控制, 新時代中國特色社會主義, 生態文明建設, 新時代中國特色社會主義思想, 法治, 全面小康社會, 香港國安法

상기 시대 키워드에 대한 TF-IDF 합산 결과는 아래와 같다.

〈표 8〉 신년사와 시대 키워드 TF-IDF

항목	JIANG	HU	XI
시대별 키워드 수	117	89	49
신년사 속 시대 키워드 수	73	46	42
TF-IDF 합산 점수	0.45056	0.20780	0.30322

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑

‘시대별 키워드 수’는 예(5)에 제시된 재임기간별 키워드 수로서 지도자 별로 각각 117개, 89개, 49개이다. ‘신년사 속 시대 키워드 수’는 신년사 내에서 어휘 및 문자 엔그램으로 구성 가능한 키워드 수이다. 장쩌민은 73개, 후진타오는 46개, 시진핑은 42개가 가능했다. 이는 ‘시대별 키워드 수’ 대비 각각 62.39%, 51.69%, 85.71%로서 시진핑의 신년사가 시대적 키워드를 가장 포괄적으로 반영했다. 그리고 ‘TF-IDF 합산 점수’는 ‘시대별 키워드’의 TF-IDF 합산값이다. 키워드 구성에 기여한 개별글자들의 TF-IDF, 즉 중요도를 합산한 결과로서 시대 키워드가 신년사에서 어느 정도의 중요도로 반영되어 있는지를 보여준다. 〈표 8〉의 결과는 각각 장쩌민 0.45056, 시진핑 0.30322, 후진타오 0.20780 순을 보이는데, 이는 장쩌민이 자신의 신년사에서 (재임기간 중의) 시대적 아젠다를 상대적으로 비중 있게 반영했다고도 볼 수 있다. 아래는 TF-IDF 상세이다.

〈표 9〉 신년사의 시대 키워드 TF-IDF 상세

순위	JIANG	TF-IDF	HU	TF-IDF	XI	TF-IDF
1	法治中國	0.02634	依法治国	0.02565	新时代中国特色社会主义思想	0.03081
2	中國加入WTO	0.02600	反分裂國家法	0.02134	新时代中国特色社会主义思想	0.02964
3	兩院三年	0.02142	國際交流	0.02134	國家安全	0.02058
4	跨越發展	0.02014	國際關係	0.02134	澳門 國情制	0.01670
5	國合項目	0.01990	國際合作	0.02134	香港 國情制	0.01670
6	國際現代化	0.01990	法治推進	0.00604	中國夢	0.01618
7	教育維新	0.01990	法治推進	0.00582	香港國法	0.01336
8	現代國防	0.01990	法治進步	0.00582	國際合作	0.01321
9	科發興國	0.01965	法治傳民	0.00539	依法治国	0.01317
10	科發興國	0.01965	港澳穩定	0.00492	國際影響力	0.01282
11	國合改革	0.01941	法治完善	0.00476	科發維新	0.01269
12	國力升級	0.01941	醫療體制	0.00455	新常態經濟	0.00989
13	國力升級	0.01941	政治改革	0.00431	创新驱动	0.00950
14	國際關係穩定	0.01941	法治建設	0.00431	科發創新	0.00950
15	國際合作	0.01941	法治政府	0.00431	新常態青	0.00872
16	國際影響力	0.01941	港澳事務	0.00431	新常態地方	0.00872
17	國際貿易	0.01941	港澳政策	0.00431	新常態	0.00872
18	高質量發展	0.01456	澳門問題	0.00431	人民為中心	0.00764
19	科發創新	0.01428	法治建設	0.00410	全面發展	0.00724
20	中英關係	0.00695	反恐安全	0.00259	全面深化改革	0.00555
21	經濟合規化	0.00433	中央軍委	0.00237	澳門事務	0.00328
22	法治進步	0.00295	中英關係	0.00237	香港事務	0.00328
23	法治深化	0.00264	社會保障	0.00200	民族復興	0.00317
24	政治體制改革	0.00225	醫療保障	0.00181	反腐敗鬥爭	0.00234
25	民族團結	0.00225	住房保障	0.00166	脫貧攻堅	0.00195
26	民族政策	0.00225	社會主義和群團社會	0.00151	軍隊現代化	0.00182
27	社會保障完善	0.00222	環境保護	0.00136	臺灣問題	0.00169
28	社會主義市場經濟	0.00215	環境保護	0.00136	香港治理	0.00167
29	港澳事務	0.00198	生態保護	0.00136	澳門穩定	0.00158
30	港澳事務穩定	0.00198	國防保護	0.00136	一帶一路	0.00156
31	法治完善	0.00191	社會穩定	0.00125	生態文明建設	0.00156
32	法治建設	0.00191	教育現代化	0.00106	共同富裕	0.00148
33	港澳穩定	0.00173	北京奧運	0.00091	反腐倡廉	0.00117
34	社會保障	0.00173	醫療改革	0.0009	文化自信	0.00117
35	社會保障完善	0.00173	和諧社會	0.00065	高質量發展	0.00093
36	社會保障改革	0.00173	和諧社會構建	0.00065	內務開放	0.00076
37	社會改革	0.00173	軍隊現代化	0.00061	臺灣政策	0.00052
38	社會穩定	0.00173	醫療體系	0.00045	臺灣統一	0.00052
39	對外開放	0.00167	醫療衛生	0.00045	二十大	0.00039
40	廉政整治	0.00108	醫療改革	0.00045	反腐敗	0.00039
41	科發進步	0.00104	教育發展	0.00045	法治	0.00039
42	香港回歸	0.00104	教育興國	0.00045	十九大	0.00026
43	香港政策	0.00104	教育改革	0.00045		
44	臺灣趨勢	0.00097	科學思想觀	0.00045		
45	臺灣關係	0.00097	住房改革	0.00030		
46	臺灣和平談判	0.00097	住房政策	0.00030		
47	臺灣政策	0.00097				
48	臺灣統一	0.00097				
49	臺灣統論	0.00097				
50	臺灣問題	0.00097				
.....				
합계		0.45056		0.20780		0.30322

‘JIANG’: 장쩌민, ‘HU’: 후진타오, ‘XI’: 시진핑

TF-IDF가 0.01이상인 키워드에는 <표 9>에서 음영처리가 되어있는데, 지도자별로는 각각 장쩌민 19개(73개 대비 26.023%), 시진핑 11개(42개 대비 26.19%), 후진타오 5개(46개 대비 10.87%) 순이다. 이들 키워드의 TF-IDF를 합산하면, 장쩌민 0.37751(0.45056 대비 83.79%), 시진핑 0.19586(0.30322 대비 64.59%), 후진타오 0.11101(0.20780 대비 53.42%) 순을 보인다. 몇 가지 두드러지는 점으로서, 첫째, 장쩌민의 상위 19개 키워드는, 총 TF-IDF의 80% 이상을 차지한다는 점에서, 신년사와의 연관성을 좌우하는 열쇠로서 기능한다. 둘째, 전체 49개 대비 42개 키워드가 신년사와 연관되어 있다는 점에서, 시진핑의 신년사는 재임기간 중의 시대 아젠다를 가장 넓게 다루고 있다. 하지만 장쩌민과 비슷한 비율의 상위 11개가 중요도 면에서는 총 TF-IDF 대비 약 65%에 그치므로 연관성에 압도적인 영향을 미친다고 보긴 어렵다. 다만 ‘新時代中國特色社會主義思想’과 ‘新時代中國特色社會主義’가 각각 0.03081과 0.02964로서 전체 키워드 중에서 가장 높은 TF-IDF를 갖고 있다는 점은 신년사를 통해 시진핑이 가장 강조하려 했던 것이 무엇인지를 가늠하게 한다.

3. 신년사와 키워드 주제

지금까지 재임기간을 중심에 두고 시대 키워드와 신년사 간 연관성을 살펴보았다. 아래는 시대 키워드를 ‘주제’별로 범주화하여 각 지도자들이 신년사에서 특별히 강조하고자 했던 주제가 무엇이었는지를 관찰해본다. 범주는 시대 키워드를 토대로 하면서 국정과제로서 언급될 만한 일곱 가지 주제로 나누어 보았다.

- [1] 국제관계와 국가위상
- [2] 정치개혁과 기강확립(핵심사상과 가치 포함), 반부패
- [3] 국방 안보와 영토, 민족문제
- [4] 경제발전과 국제무역

- [5] 과학기술과 인재교육
- [6] 빈곤퇴치와 민생개선, 사회안정
- [7] 의료체제와 복지, 삶의 질

그리고 이들 범주 안으로 <표 9>의 키워드를 재배열하였다. <표 10>은 그 결과로서 주제별 키워드와 그것의 TF-IDF를 보여준다('[1], [2] ...'는 상기 제시된 주제의 번호를 가리킨다).

<표 10> 신년사와 주제별 키워드 TF-IDF

범주	JIANG	HU	XI
[1]	中國加入WTO, 跨國投資, 國際關係穩定, 國際合作, 國際影響力, 國際貿易, 中美關係, 經濟全球化, 對外開放, 對外經濟合作, 海外華人, 東西部開發, 貿易開放, 北京申奧 [TF-IDF: 0.14054]	國際交流, 國際關係, 國際合作, 中美關係, 北京奧運 [TF-IDF: 0.06730]	國際合作, 國際影響力, 一帶一路, 內外開放 [TF-IDF: 0.02835]
[2]	法治中國, 兩個一百年, 法制進步, 法制深化, 政治體制改革, 法制完善, 法制建設, 社會改革, 腐敗整治, 改革開放, 三個代表, 民主法治, 腐敗治理, 民主建設 [TF-IDF: 0.06455]	依法治國, 法治推進, 法制進步, 法治進展, 法治完善, 政治改革, 法治建設, 法治政府, 法制建設, 中央軍委, 社會主義和諧社會, 和諧社會, 和諧社會構建, 科學發展觀 [TF-IDF: 0.07614]	新時代中國特色社會主義思想, 新時代中國特色社會主義, 中國夢, 依法治國, 人民為中心, 全面深化改革, 民族復興, 反腐敗鬥爭, 反腐倡廉, 文化自信, 二十大, 反腐敗, 法治, 十九大 [TF-IDF: 0.11227]
[3]	國防現代化, 現代國防, 國防升級, 國防科技, 民族團結, 民族政策, 港澳事務, 港澳事務穩定, 港澳穩定, 香港回歸, 香港政策, 臺海局勢, 臺灣關係, 臺灣和平談判, 臺灣政策, 臺灣統一, 臺灣統一論, 臺灣問題, 對臺政策, 澳門回歸, 澳門政策, 軍隊現代化 [TF-IDF: 0.10052]	反分裂國家法, 港澳穩定, 港澳事務, 港澳政策, 港澳問題, 反恐安全, 軍隊現代化 [TF-IDF: 0.04239]	國家安全, 澳門一國兩制, 香港一國兩制, 香港國安法, 澳門事務, 香港事務, 軍隊現代化, 臺灣問題, 香港治理, 澳門穩定, 臺灣政策, 臺灣統一 [TF-IDF: 0.06170]
[4]	國企重組, 國企改革, 社會主義市場經濟, 經濟結構調整, 經濟調整, 發展開放區, 經濟快速發展, 經濟快速增長, 現代化建設 [TF-IDF: 0.04548]	[TF-IDF: 0.00000]	新常態經濟, 高質量發展 [TF-IDF: 0.01082]
[5]	教育強國, 科技興國, 科教興國, 高新技術發展, 科技創新, 科技進步, 高校建設, 教育現代化 [TF-IDF: 0.09009]	教育現代化, 教育發展, 教育振興, 教育改革 [TF-IDF: 0.00241]	科技強國, 創新驅動, 科技創新 [TF-IDF: 0.03169]
[6]	社會穩定, 農村振興 [TF-IDF: 0.00197]	社會保障, 住房保障, 社會穩定, 住房改革, 住房政策 [TF-IDF: 0.00551]	新冠疫情, 新冠疫情防控, 全面小康社會, 脫貧攻堅, 共同富裕 [TF-IDF: 0.02811]

범주	JIANG	HU	XI
[7]	社會保障強化, 社會保障, 社會保障完善, 社會保障改革 [TF-IDF: 0.00741]	醫療體制, 醫療保障, 環保政策, 環境保護, 生態保護, 資源保護, 醫療改善, 醫療體系, 醫療衛生, 醫療改革 [TF-IDF: 0.01405]	新能源, 生態文明建設 [TF-IDF: 0.01028]

'JIANG': 장쩌민, 'HU': 후진타오, 'XI': 시진핑

〈표 10〉에서는 TF-IDF합산값을 내림차순에 따라 음영으로 표시했다 (TF-IDF가 높으면 음영이 짙고 낮으면 옅다). 장쩌민이 신년사에서 가장 중요시했던 주제는 ‘[1]국제관계와 국가위상’, ‘[3]국방 안보와 영토, 민족문제’, ‘[5]과학기술과 인재교육’ 순이다. 후진타오는 ‘[2] 정치개혁과 기강 확립(핵심사상과 가치 포함), 반부패’, ‘[1] 국제관계와 국가위상’, ‘[3] 국방 안보와 영토, 민족문제’ 순이고 시진핑은 ‘[2] 정치개혁과 기강확립(핵심사상과 가치 포함), 반부패’, ‘[3] 국방 안보와 영토, 민족문제’, ‘[5] 과학기술과 인재교육’ 순으로 중요시했다.

장쩌민은, [국제/국가위상(0.14054)] 대 [국방/영토/민족(0.10052), 과학/교육(0.09009)]으로 대별될 정도로, 신년사에서 ‘국제관계와 국가위상’을 높은 비중으로 강조했다. WTO가입, 다자주의와 다자외교, 올림픽 개최 확정 등 경제성장에 토대를 둔 ‘국제관계와 국가위상’의 문제가 장쩌민에게 가장 중요한 과제였던 것으로 보인다. 그리고 홍콩과 마카오 반환, 대만통일 등의 ‘국방, 영토, 민족문제’와 고등교육 현대화를 위한 211공정과 985공정 등의 ‘[5]과학기술, 인재교육’문제가 그다음 시대 과제로서 중요시되었다.

후진타오는 [정치/사상/가치(0.07614), 국제/국가위상(0.06730)] 대 [국방/영토/민족(0.04239)]의 대별을 보이며 ‘정치사상과 이념’, ‘국제관계와 국가위상’을 강조하고 있으나 다른 두 지도자에 비해 TF-IDF가 상대적으로 부각되어 보이지는 않는다. 대외적으로는 국제적 위상이 제고됐으나 도광양희(韜光養晦) 기초의 외교를 유지하려 했고, 대내적으로는 경제불평등과 사회불평등 해결을 시대과제로 받아안아 ‘和諧社會’ 기초의 사회 안정

을 피하는데 치중했던 결과로 해석된다. 이 점은 장쩌민보다 후진타오 재임기간 동안 ‘[6]빈곤퇴치와 민생개선, 사회안정’, ‘[7]의료체제와 복지, 삶의 질’ 관련 키워드가 늘어나 있고 관련 키워드의 TF-IDF 또한 장쩌민에 비해 높다는 점에서 확인된다.

시진핑은 [정치/사상/가치(0.11227)] 대 [국방/영토/민족(0.08170), 과학/교육(0.03169)]의 대별로써 ‘정치사상과 이념, 정치개혁과 기강확립, 반부패’를 상대적으로 중요시했던 것으로 보인다. 시진핑 시기부터는 일대일로(一帶一路)에 의한 국제사회 역할 모색, 미국과의 강대강 대치 등 전임 지도자들의 도광양회 기조를 벗어남으로써 ‘국제관계와 국가위상’ 문제가 시대의 우선과제로서 대두되지 않는 특징을 보인다. 0.11227의 TF-IDF합산 값이 보여주듯, 시진핑의 신년사에서는 ‘新時代中國特色社會主義思想’에 기초한 법치(‘法治’), 개혁심화(‘深化改革’), 기강확립(‘反腐敗’), 그리고 이를 통한 당중심의 단결과 민족부흥(‘中國夢’) 등이 특별히 강조되어 있다.²⁰⁾ 이 같은 민족의 단결과 부흥의 기치는, 신년사에서 두 번째로 강조된, 홍콩과 마카오 등에 관한 ‘국방 안보와 영토, 민족문제’로 이어진다고 볼 수 있다.

V. 결론

지금까지 장쩌민, 후진타오, 시진핑의 30년 신년사를 크게 TTR(어휘다양도)과 TF-IDF(어휘빈도-역문서빈도)의 관점에서 분석하였다. 신년사 전체 TTR은 0.196으로서 반복의 경향이 높은 것으로 나타났다. 모두에게 메시지가 전달될 수 있도록 해야 하는 대국민연설로서 신년사가 기능하기 때문이다. 다만, 지도자별로는 큰 차이를 보였는데, 시진핑의 TTR은

20) Wikiwand(習近平新時代中國特色社會主義思想): <https://www.wikiwand.com/zh-cn/articles/習近平新時代中國特色社會主義思想>.

0.407로서 어휘다양도 면에서 장쩌민(0.224), 후진타오(0.180)보다 약 2배가 높다. 신년사 당 평균문장수도 시진핑이 가장 많았던(42개) 것을 볼 때 (상대적으로) 많은 정보량과 다양한 어휘 사용을 시진핑의 언어 특징으로 볼 수 있을 것이다. 그에 반해, 후진타오는 고빈도 상위 100개 어휘가 총 내용어 대비 56.59%를 차지하며 시진핑(32.62%)과 2배 가까운 차이를 보였다. 이는 후진타오가 시진핑과의 대척점에서 소수의 고빈도 어휘를 반복해서 사용하고 있음을 알 수 있다.

TF-IDF 면에서 세 지도자의 공통 핵심어는 ‘中國’, ‘人民’, ‘發展’, ‘世界’를 꼽을 수 있다. 이들 핵심어는 지위가 30년간 변함없었음에도 그에 대한 TF-IDF는 시진핑 신년사부터 급격히 낮아졌다. IDF 측면에서의 지도자별 고유핵심어는 장쩌민 1개, 후진타오 2개, 시진핑 10개인데 이는, TF의 양상과 반대로, 시진핑부터 급격히 높아졌다. 시대 키워드와 신년사 간 관계에서, 포괄성은 시진핑이 가장 높았고 연관성은 장쩌민(0.45056)의 TF-IDF가 시진핑 대비 0.5배, 후진타오 대비 2배 이상 높았다.

이상의 분석 결과는 다음 두 가지 함축을 갖는다. 첫째, 3인의 신년사에는 ‘[장쩌민, 후진타오] 대 [시진핑]’의 구도가 일관되게 유지된다. 그리고 그 사이에는 ‘반복’과 ‘다양함’, ‘관례’와 ‘차별화’라는 대별이 존재한다. 장쩌민과 후진타오는 ‘반복’의 경향이 강한 반면 시진핑은 ‘다양’한 어휘사용 경향이 강하다. 특히 후진타오와 시진핑 간 대별은 명확하다. 더불어 장쩌민과 후진타오의 언어사용은 ‘관례’적이다. 신년사에서 장쩌민, 후진타오는 공통 핵심어를 일관되게 중시했고 자신만의 고유핵심어도 부각시키지 않았다. 반면에 시진핑은 공통 핵심어에 동일 지위를 부여하고는 있으나 TF-IDF가 전임 두 지도자에 비해 크게 낮았고 그에 반해 자신만의 핵심어는 부각시켰다. 다만 TF-IDF가 낮다고 하여 ‘중요시되지 않았다’고 보는 데는 재고가 필요하다. 어휘를 다른 두 지도자보다 다양하게 사용하고 (TTR: 0.407) 시대 키워드에 대한 포괄성 역시 높다(85.71%)는 점에서 시진핑의 언어사용은 전임 두 지도자와 비교해 ‘차별화’ 경향을 갖는다고 보는 것이 합리적일 것이다.

둘째, 시대 키워드를 범주화하여 봤을 때, 신년사에서 중요시했던 장쩌민의 주제는 ‘국제’로 대표된다. 도광양회의 태도를 견지하면서, 국가위상의 제고와 어느 한 국가로의 쏠림없는 다자외교를 중시해야 했던 시대 과제의 반영으로 볼 수 있다. 그에 반해 후진타오는 ‘민생’으로 대표된다. 후진타오도 장쩌민과 비슷하게 ‘정치-국제-국방/민족’ 등의 주제를 중요시하지만 장쩌민과 비교해 ‘빈곤’과 ‘의료’ 부문 키워드가 크게 늘어나 있는 점에서 후진타오가 중요시한 주제를 ‘민생’으로 보는 것이 합리적이라고 본다. 마지막으로 시진핑이 중요시했던 주제는 ‘정치’로 대표된다. 이는 ‘시진핑 시대, 중국만의 사회주의(新時代中國特色社會主義)’를 강조하며 국가기강을 재확립하고 내부 결속을 다져야 했던 시대 과제의 반영으로 해석할 수 있을 것이다.

< 참고문헌 >

- 김가현, 김지영, <절대평가 도입 전·후 수험능력시험 영어 읽기 지문의 어휘 분석 비교>, 《외국학연구》, 제58집, 2021.
- 박찬욱, <중국국가지도자 신년사의 행위구조와 ‘我們’—江澤民的 1997년, 1999년 신년사를 중심으로>, 《중국어문학논집》, 제57호, 2009.
- 오힘찬, 《제미나이 노트북LM 미친활용법 51제》, 서울: 골든래빗, 2025.
- 이정진, 하두진, 박민준, <텍스트 마이닝을 통해 본 시진핑 시기 중국의 전략: 후진타오 시기와의 비교를 중심으로>, 《인문사회과학 연구》, 제31권 제4호, 2023.
- 이지원, <시진핑 정치 담화에서 나타나는 ‘我們’—비판적 담화분석의 관점을 중심으로>, 《중국문학연구》, 제79집, 2020.
- 젠스 알브레히트, 싯다르트 라마찬드란, 크리스티안 뵵클러 저, 심상진 역, 《과이썬 라이브러리를 활용한 텍스트 분석》, 서울: 한빛미디어, 2022.

Hope, Arran, “Dreams Deferred in Xi’s New Year’s Speech”, *China Brief* 25, no.1, 2025.

Mikheev, Andrei, “Text Segmentation”, in Mitkov, Ruslan ed., *The Oxford Handbook of Computational Linguistics* (2nd edition), Oxford University Press, 2022.

百度百科(1992年~2025年新年賀詞): https://baike.baidu.com/starmap/view?nodeId=dfe1324092db5f68bfb97f8b&lemmaTitle=1997%E5%B9%B4%E6%96%B0%E5%B9%B4%E8%B4%BA%E8%AF%8D&lemmaId=24241066&starMapFrom=lemma_starMap&fromModule=lemma_starMap.

大公網(元旦賀詞): <http://news.takungpao.com/special/yuandanheci/>.

人民日報: <https://reader.jojokanbao.cn/rmr/>.

人民網(習近平系列重要講話數據庫): <http://jhsjk.people.cn/>.

Gemini: <https://gemini.google.com/app>.

GitHub(goto456/stopwords): <https://github.com/goto456/stopwords>.

HanLP: <https://pypi.org/project/hanlp/>.

Perplexity: <https://www.perplexity.ai/>.

Stanza: <https://stanfordnlp.github.io/stanza/>.

Wikiwand(科學發展觀): <https://www.wikiwand.com/zh/articles/科學發展觀>.

Wikiwand(和諧社會): <https://www.wikiwand.com/zh/articles/和諧社會>.

Wikiwand(習近平新時代中國特色社會主義思想): <https://www.wikiwand.com/zh-cn/articles/習近平新時代中國特色社會主義思想>.

< Abstract >

The purpose of this study is to quantitatively examine the extent to which the language of Chinese leaders reflects their respective historical contexts by analyzing the New Year's addresses delivered by Jiang Zemin, Hu Jintao, and Xi Jinping from 1993 to 2022. To this end, the study performs a comprehensive linguistic analysis of thirty years of speeches, utilizing word frequency and Type-Token Ratio (TTR) analysis, TF-IDF-based keyword extraction, and era-specific keyword association analysis. The results, grounded in TTR and TF-IDF metrics, suggest that while Jiang and Hu tended to exhibit a conventional linguistic pattern characterized by repetition and formality, Xi appears to have adopted a more diverse and differentiated linguistic strategy. Furthermore, regarding era-specific keywords, the analysis indicates a shift in priorities: Jiang placed a relatively higher emphasis on elevating China's international status, Hu focused more on livelihood improvements for social harmony, while Xi seems to have prioritized political values to advance socialism with Chinese characteristics.

Key Words : 중국지도자(Chinese leader), 신년사(New Year's Address), 어휘빈도(Word Frequency), 어휘다양도(TTR), 어휘빈도-역문서빈도(TF-IDF)