

## 중국 학술지 ‘韓國語教學与研究(한국어교수와연구)’의 연구 동향 분석

왕서기 \* 김유미 \*\*

### -차 례-

1. 들어가는 말
2. 연구 방법
  - 2.1 자료 수집 및 분석 절차
  - 2.2 측정 도구
3. 연구 결과
  - 3.1 논문 수 추이
  - 3.2 언어 네트워크 분석 결과
  - 3.3 토픽 모델링 분석 결과
4. 결론 및 제언

---

\* 서울과학종합대학원대학교 한국교육경영학 박사과정 수료.

\*\* 교신저자, 서울과학종합대학원대학교 한국교육경영학 교수.

## [국문초록]

본 연구는 2016년부터 2023년까지 중국 학술지 '韩国语教学与研究(한국어 교수와 연구)'에 게재된 논문들의 언어 사용 및 연구 주제의 변화를 분석하고 이에 대한 한국어 교육적 시사점을 제언하는 데 그 목적이 있다. 분석 대상 논문은 총 711편으로 이 중에서 중국어로 발표된 272편은 복수의 번역자가 신뢰도 검정을 거쳐 번역하였다. 자료 수집은 중국 국가 지식 인프라(CNKI) 사이트를 통해 논문을 수집하고, 전처리 과정을 거쳐 데이터 분석을 수행하였다. 측정도구는 키워드 빈도 분석, 언어 네트워크 분석, 시계열 회귀분석 및 토픽 모델링을 통해 연구의 주요 주제를 도출하고 시간의 흐름에 따른 변화 양상을 파악하였다.

첫째, 언어 네트워크 분석 결과, 키워드 출현 빈도분석과 중심성이 높은 어휘는 '번역, 교육, 언어, 문학' 등이 핵심 주제어로 나타났다. 이를 통해 해당 주제와 관련 연구가 학술지에서 중요한 위치를 차지하고 있으며, 이들 키워드를 중심으로 연구 주제들의 군집화가 이루어졌음을 확인할 수 있었다. 둘째, 토픽 모델링을 통해서 총 9개의 주요 연구 주제가 도출되었으며, 이 중 교육 관련 주제가 전체의 가장 높은 비율을 차지했다. 셋째, 연구 주제의 변화 양상을 시계열 분석한 결과, 문학 관련 연구는 시간이 지남에 따라 감소세가 명확했으며 교수, 전공/대학 관련 연구는 증가 추세를 보였다.

본 연구는 중국 학술지에서 한국어 교육 연구의 학문적 관심이 변화하고 있음을 보여주며, 특히 중국어 논문의 비율이 증가하고 있다는 점에서 중요한 시사점을 제공한다. 향후 연구에서는 이와 같은 변화가 지속될 가능성을 모니터링하고, 더 정교한 분석 도구를 활용해 장기적인 연구를 진행할 필요가 있다.

**주제어** : 중국에서의 한국어 교육, 연구 동향 분석, 시계열 회귀분석, 언어 네트워크 분석, 토픽 모델링

## 1. 들어가는 말

본 연구는 과학적 기법을 사용하여 중국 학술지 ‘韩国语教学与研究(한국어 교수와 연구)<sup>1)</sup>에 게재된 논문들의 주요 연구 주제의 변화를 점검하고, 향후 중국에서 ‘한국어 교육’ 연구의 방향성에 대해서 시사점을 제시하는 데 그 목적이 있다. 이를 위해서 토픽 모델링(topic modeling)과 언어 네트워크 분석(semantic network analysis) 기법을 활용하여 총 711편의 논문을 수집하고 이를 분석하고자 한다. 이 중에서 272편의 논문이 중국어로 쓴 논문이어서 한국어로 번역하여 분석하였다.

1992년 중한 수교 이후 각 분야에서 활발한 교류와 협력이 이뤄졌고, 한국어는 양국 문화교류의 중요한 매개체로서 중국에서도 발전했다. 또한 이러한 교류와 협력은 한국어 교육의 학문적 발전을 증가시켰다. 중국에서 ‘한국어 교육’과 관련된 연구 주제로는 이론적인 한국어 문법 연구, 한국어 교육 방법, 한국어 교육 연구 방향의 제언 등과 같은 순서로 연구하고 있다(王菲, 2022; 金银花, 2023; 张晶·王平平, 2019).

이에 중국에서의 ‘한국어 교육’ 관련 연구 동향을 분석한 논의들을 찾아볼 수 있다. 먼저 张晶·王平平(2019)은 2014~2018년 중국 사이트에 발표된 562편의 석박사 논문을 대상으로 각각 논문의 시기, 연구 유형, 연구 주제로 구분하고 연구 동향을 분석하였다. 연구의 주제 분류에서는 연구자가 언어, 문학, 문화, 번역, 교육 다섯 가지로 나누어 분석하였다. 이 중에 문학과 문화 하위 분류에서 모두 ‘사상 연구’ 주제를 포함하고 있어서 연구자 개인의 주관적 분석을 벗어난 과학적인 분석 방법을 통한 객관적 분류 방법의 필요성이 제기된다. 최근 정보통신 기술의 발달로 대규모 데이터 처리 방법을 언어 연구에서도 응용하고 있다. 데이터 분석 기법은 데이터의 양과 복잡성이 증가함에 따라 진화해 왔다. 초기의 소규모 자료를 대상으로 한 통계적 분석 기법에서부터 현재의 빅데이터 분석을 위한 AI 분석에 이르기까지, 데이터 분석의 범위와 능력은 계속해서 확장·진화하고 있다. 특히 자연어 텍스트 내에 포함된 주제나 관계를 발견하기 위한 분석 기법으로 토픽 모델링(Topic

1) 본 논문에서 중국 학술지 ‘韩国语教学与研究’를 중국어가 아닌 ‘한국어 교수와 연구’로 한국어로 기술하겠다.

Modeling)이 사용되고 있다(Blei D. M., 2012).

하지만 중국에서 빅데이터 분석 기법을 활용한 한국어 교육의 연구 동향을 분석한 연구는 많지 않다. 薛娇(2023)는 2000년부터 2021년까지 중국 학술지에 발표한 연구를 한국 KCI 데이터베이스에서 검색하여 539개 문헌을 연구 대상으로 분석하였다. NetMiner 4.0 소프트웨어의 시각적 분석과 LDA 모델링 기능을 사용하여 20년 동안 중국어 원어민의 한국어 교육에 관한 연구 현황을 분류하였다. 그러나 중국 본토에서 발표한 ‘한국어 교육’의 연구 동향 분석은 아직 활발하게 이루어지지 않고 있으며 이에 연구의 필요성이 제기된다.

한국에서는 중국인 학습자에 관한 한국어 교육 연구가 활발하게 이루어지고 있고, 한국어 교육의 연구 동향을 점검하는 논의도 많은 관심을 나타내고 있다. 이정희(2014)의 연구는 학술연구정보서비스(RISS)와 국회도서관에서 1,288편의 연구를 수집하고 중국인 학습자 대상으로 한국어 교육 동향을 연도별, 유형별, 주제별, 연구 방법별로 살펴보았다. 그러나 분류 기준이 명료하지 않아 연구자가 이론적 분류 기준을 보완하여 재구성하였지만, 상위 주제 ‘학습자 요인’과 ‘학습자 사전 개발’은 공통적인 학습자 요인으로 볼 수 있으므로 중복된 부분이 있다. 원미진(2018)에서는 중국어권 학습자의 한국어 교육 연구 동향을 분석하기 위해 RISS에서 1,444편 석박사 학위논문을 수집했다. 논문 수집 과정은 연구자가 검색어를 통해 직관적으로 문헌을 수집했기에 누락된 자료도 있을 것으로 예측된다. 연구 주제 분류도 연구자가 문법, 어휘, 발음, 담화/화행, 기타로 분류하여 객관적이고 과학적인 방법이 부족하다. 김민희(2020)는 중국인 학습자 대상으로 한국어 습득 연구를 LDA 분석 기법으로 사용했지만, 연구 대상이 2003년부터 2020년까지의 한국 국내 학위논문과 학술지 논문만 분석을 하였다.

이상의 연구들은 연구 대상이 중국에서 발표한 연구 결과물이 아니다. 조숙근·김설연(2022)은 중국인 대상으로 RISS에서 483편의 학술지 논문들을 수집하고 한국어 교육의 연구 동향을 분석하였다. 연구자가 주관적으로 설정한 분류 기준으로 선행연구를 분석했지만, 상위 주제 ‘한국어 교수·학습’에 하위 주제가 ‘기능별 교수’가 나온다. 이에 대한 분석 결과를 보면 ‘기능별 교수’에 문법(25편), 어휘(18편), 발음(14

편), 문학(9편), 읽기(6편), 말하기(3편), 듣기(2편)로 나타나는데 이는 상위 주제 ‘한국어 교육 내용’과 중복된 부분이 많이 있어 연구이 한계점이라고 할 수 있겠다.

이상의 선행연구를 종합해 본 결과, 중국 내 한국어 교육 관련 학술지를 대상으로 과학적 기법을 사용하여 연구의 흐름을 점검하고, 향후 교육의 방향성에 관해서 제언하는 연구는 지금까지 많지 않다. 이러한 선행연구의 한계점을 보완하기 위하여 본고에서는 중국 내 학술지 ‘한국어 교수와 연구’에 등재된 논문 711편을 수집하여 언어 네트워크 분석 기법과 토픽 모델링 기법을 활용한 한국어 교육의 연구 동향을 살펴보고 이에 대한 교육적 시사점을 제시하고자 한다.

## 2. 연구 방법

### 2.1 자료 수집 및 분석 절차

분석 대상은 중국 내 한국어 교육 학술지 ‘한국어 교수와 연구’로 2016~2023년까지 게재된 711편의 논문을 대상으로 연구 제목의 정보를 수집하였다. 자료 수집 및 분석 절차를 정리하면 다음과 같다.

첫 번째 과정으로, 중국 최대의 논문 검색 사이트인 중국 국가 지식 인프라(中国知网 : CNKI)<sup>2)</sup> 해외 사이트를 통해서 ‘한국어’, ‘한국어 교육’, ‘한국문화’, ‘조선어’, ‘조선어 교육’ 등의 검색어를 입력한 후 관련 논문들을 확인하였다. 검색된 논문 중에서 학위 논문을 제외한 학술지를 기준으로 재차 분류하고 서지정보를 엑셀로 저장하였다. 한국어 교육과 관련 있는 학술지는 ‘중국조선어문(中国朝鲜语文 : Korean Language in China)<sup>3)</sup>과 ‘한국어 교수와 연구(韩国语教学与研究 : Korean Teaching and

---

2) 국가 지식 인프라(National Knowledge Infrastructure, NKI)는 중국에서 가장 큰 논문 검색 사이트다. NKI는 중국 국내에서 검색 가능한 사이트와 해외에서 검색 가능한 사이트로 나뉘며, 이 두 사이트는 검색 범위에서 차이를 보인다. 해외에서 검색하는 경우 일부 논문의 수집이 제한된다.

3) ‘중국조선어문(中国朝鲜语文, Korean Language in China)’은 중국에서 조선어(한

Research)’<sup>4)</sup>이다. 이들 학술지에 등재된 논문을 연도를 기준으로 수집·정리하였다. ‘중국조선어문’의 경우 모든 논문이 한국어로 되어 있었기 때문에 별도의 번역 과정 없이 수행할 수 있었다. 이와 관련된 분석 결과는 왕서기·김유미(2024)를 통하여 발표하였다.

본고에서는 ‘한국어 교수와 연구’ 학술지 272편의 논문을 대상으로 연구 동향 분석을 시행한다. 이 학술지는 중국어로 작성되어 있어서 이를 분석하기 위해 별도의 한국어 번역 과정이 필요하다. 그래서 두 번째 과정에서는 중국어 원어민 2명<sup>5)</sup>이 번역에 참여하여 제목과 초록을 번역하였다. 번역 과정은 각각의 번역자가 스스로 2회 번역(1차, 2차)을 하고 평가자 내 신뢰도<sup>6)</sup>를 측정한 후에 번역물을 상호비교하여 수정하였다. 그 뒤 번역물에 대한 상호 검증을 수행한 후 다음으로 자기 조정을 거쳐 최종 번역물을 확정하였다.

두 명의 번역자가 번역한 최초 결과물에 차이가 발생하였는데 <표 1>과 같다. 이 부분은 의견을 조정해서 하나의 용어로 통일하고 작업하였다.

---

국어) 관련 학술 연구를 다루는 중요한 학술지 중 하나이다. 이 학술지는 1977년에 창간되어 중국조선어문 잡지사가 주관하며 발행하고 있다. 주요 연구 주제는 조선어의 언어학적 분석, 교육 방법론, 문화적 연구 등 한국어와 관련된 다양한 분야를 포함한다(왕서기·김유미, 2024). 이 학술지는 중앙민족대학을 포함한 약 10개 대학에서 핵심 학술지로 지정되어 있으며, 이는 중국 국내 학계에서 높은 신뢰도와 권위를 가지고 있음을 의미한다. 2019년에 ‘중국조선어문’은 중국 국가 지식 인프라(National Knowledge Infrastructure, NKI)에 공식적으로 등재되었으며, 이에 따라 2019년 이후의 정보만 온라인에서 검색 및 접근할 수 있다.

- 4) ‘한국어 교수와 연구(韩国语教学与研究, Korean Teaching and Research)’는 중국에서 한국어 교육 및 연구와 관련된 유일한 전문 학술지라고 할 수 있다. 이 학술지는 2012년에 창간되었으며, 헤이룽장(黑龍江)성에서 발행되고 있다. 한국어 교육 및 연구에 대한 최신 학술 성과와 연구 동향을 다루며, 한국어 교사 및 학자들에게 학술 토론과 연구 플랫폼을 제공한다(왕서기·김유미, 2024).
- 5) 번역에 참여한 2인은 모두 한국어 교육 전공자이며 1명은 중국에서 한국어 교육 박사 과정생이며 나머지 1명은 중국에서 한국어 교사로 활동 중이다.
- 6) 번역자가 2회 번역한 번역물의 신뢰도 분석 결과는 번역자 1이 99%였고 번역자 2가 95%였다. 번역자 간 신뢰도는 97.79%로 나타났다.

<표 1> 번역이 다른 단어들

	논문 제목 중 어휘 (중국어)	편수	최종 번역 (한국어)
1	教学	41	교수
2	译介	3	번역 및 수용
3	被动语态	2	피동
4	误读	1	오독

세 번째 과정으로, 논문 제목에서 핵심 키워드를 추출하기 위해서 데이터 정제하는 전처리 과정을 거쳤다. 이를 위해서 유의어, 지정어, 제외어를 사전으로 구성하였다. 먼저 유의어의 구성은 학술 논문 저자들이 각각 다른 단어를 사용하여 표기했지만 사실상 유사한 의미를 가지는 키워드를 단일한 용어로 통합하는 과정을 말한다. 예를 들어서 ‘중국 전문대학, 중국 전문대학교, 중국 대학, 중국대학’ 등은 모두 동일한 개념을 나타내므로 ‘중국대학’으로 수렴할 수 있도록 유의어 사전에 등록하였다. 다음으로 지정어 사건의 구성은 형태소 분석 과정 중에 의미가 분리되지 않도록 특정 용어를 정의하는 사전이다. 대표적인 예로 ‘교육과정’은 2개의 별개의 단어인 ‘교육’과 ‘과정’으로 구성되어 있는데 각각의 단어로 분석되지 않도록 하나의 단어로 등록하였다. 끝으로 제외어 사전은 분석에서 제외해야 할 단어들을 명시하고 분석 대상에 포함하지 않았다. ‘한국’, ‘연구’, ‘중심’ 등과 같이 연구 주제의 특징을 명확히 드러내지 않으면서 전체 텍스트 내에서 출현 빈도가 높아 상대적으로 그 중요성이 낮은 단어들이다. 마지막으로 더 정확한 주제어를 추출하기 위해 무의미한 1음절 단어 ‘ㄱ7’, ‘문’, ‘속’ 등도 제외어 사전에 포함하였다.

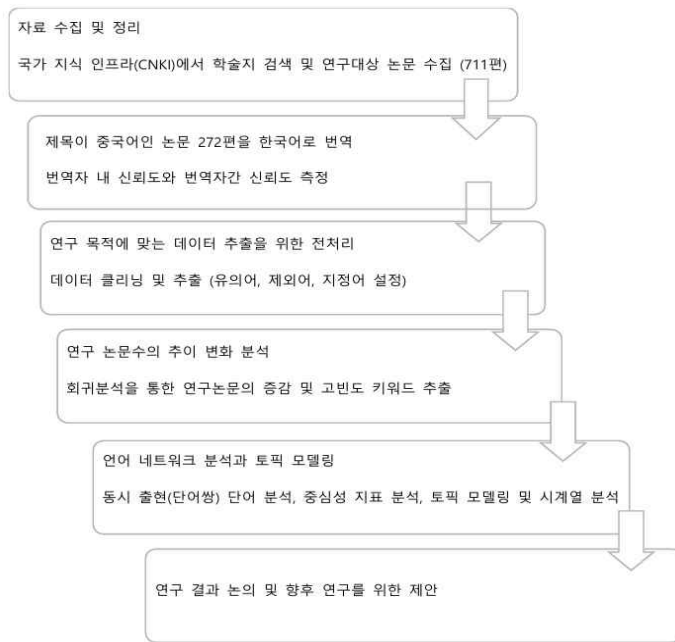
네 번째 과정은, 자료를 연도별로 정리하고 논문 개수의 변화를 분석하였다. 먼저 연도별로 한국어와 중국어 논문 구성 비율을 살펴보았다. 각각의 언어로 발표된 논문의 증감 추이를 명확하게 살펴보기 위해 시계열 회귀분석을 실행하였다. 시계열 회귀분석은 모두 세 구간으로 나

7) 예를 들면, ‘한국어 역행 유음화와 “ㄴ+ㄹ” 치조비음화의 관련성 연구’와 같은 논문 제목에서 “ㄴ”, “ㄹ”의 경우이다.

누어 분석하였는데 먼저 전체 구간인 2016년부터 2023년까지를 분석하였고, 다음으로 논문 발표에 영향을 주었을 가능성이 높은 외생 변수의 발생 시점인 코로나 발생 연도(2020년)를 기준으로 그 이전인 2016년부터 2020년까지, 코로나 이후인 2021년부터 2023년까지를 분석하였다. 또한, 전처리가 된 자료를 바탕으로 빈도분석을 수행하여 핵심 키워드를 추출하였다. 이를 통해 고빈도 단어를 순서대로 정리하고, 이를 표와 그림으로 시각화하여 제시했다.

다섯 번째로, 네트워크 분석을 통해 동시에 출현하는 단어 및 중심성 지표를 도출하고 이를 비교하였다. 특히, 연결 중심성을 중심으로 핵심 단어 간의 관계를 살펴보고 이를 시각화하였다. 네트워크 분석 결과를 기반으로 토픽 모델링 과정을 진행하여 전체 텍스트를 구성하는 연구 주제와 각 연구 주제와 관련된 핵심 단어를 확인했다. 또한, 연도별 연구 주제의 비율 변화를 시계열 회귀분석을 통해 분석하고, 이러한 분석 결과를 표와 그림으로 시각화했다.

마지막으로 연구의 결과를 정리하고, 본 연구의 한계점과 향후 연구를 위한 제안 사항을 제시하였다. 지금까지의 과정을 정리하면 <그림 1>과 같다.



<그림 1> 연구 절차

## 2.1 측정 도구

### 2.2.1 언어 네트워크 분석(Language Network Analysis)

언어 네트워크 분석은 텍스트에서 의미를 가지는 단어들을 추출하고, 이들 간의 연결 관계를 파악하여 네트워크를 생성하고 분석하는 방법이다(이수상, 2014). 이 분석은 단어와 단어들 사이의 관계를 통해서 텍스트에 내재하고 있는 구조적 특성을 분석하며, 중심성 지표를 활용해 핵심 단어의 중요도를 평가한다. 연결 중심성, 근접 중심성, 매개 중심성이 주로 사용되며, 각 단어의 영향력을 분석한다(신우봉·김일환·김흥규, 2010; 최일영·이용성·김재경, 2010).

분석 과정은 문서나 게시물 등 텍스트 자료를 수집하고 전처리 과정을 통해 정제하며, 불용어 제거와 형태소 분석 등을 수행하여 데이터를 준비한다. 다음 단계에서는 단어 또는 구문 간의 빈도나 동시 발생 등

의 관계를 기반으로 네트워크를 구축한다. 이 네트워크를 바탕으로 중심성, 클러스터링, 연결성 등을 분석하여 언어적 패턴과 의미를 도출하며, 마지막으로 분석 결과를 시각적으로 표현하여 그 관계를 살펴본다(Christensen, A. P. & Kenett, Y. N., 2021; 김동아, 2023).

언어 네트워크 분석은 키워드 간의 동시 출현 관계를 중점적으로 분석하고 네트워크 시각화 도구를 통해 분석 결과를 직관적으로 전달할 수 있는 장점이 있어서 텍스트의 의미적 구조를 파악하는 데 유용하다. 이점은 연구자의 주관적 판단을 넘어서 연구의 신뢰성과 타당성을 높이는데 도움이 된다(이수상, 2014).

본 연구에서도 빈도분석, 동시 출현 단어 분석, 중심성 분석을 수행하였으며 이를 키워드 네트워크 수형도로 시각화하였다.

### 2.2.2 토픽 모델링(Topic Modeling)

토픽 모델링은 비정형 텍스트 데이터를 분석하여 텍스트 내의 주제를 자동으로 추출하는 기법이다. 이를 활용하면 대규모 텍스트 데이터를 효율적으로 분류하고 중요한 정보를 식별할 수 있다. 예를 들어, 정치적 여론 분석, 소비자 댓글 분석, 학술 문헌 탐색 등에서 사람들이 관심사가 무엇인지 확인할 수 있다.

정확한 토픽 모델링을 위해서 그동안 다양한 알고리즘<sup>8)</sup>이 개발되었고 현재도 개선되고 있지만, 연구자들이 가장 많이 사용하는 것은 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA)이다. LDA는 Blei 외(2003)에 의해 제안된 주제 추출 기법으로 주제를 구성하는 특정 단어가 확률 분포를 가지고 더 자주 나타난다는 가정을 바탕으로 문서 속에 내포되어 있는 여러 주제를 추출한다. LDA는 여러 가지 장점이 있는데 특히 대규모의 텍스트 데이터에서 잠재적인 주제를 자동으로 추출하는 데 매우 효과적이며 새로운 문서가 추가될 때도 기존의 주제 구조를 유지하면서 빠르게 적응할 수 있으며 명료한 해석이 가능하다

8) 알고리즘이란 특정 작업을 수행하기 위해 명확하게 정의된 계산 절차로 특정한 계산 문제를 해결하기 위한 도구이다(Cormen, Leiserson, Rivest, & Stein, 2009:5).

(Vayansky, I., & Kumar, S. A, 2020).

한국어 교육 분야에서도 연구 동향 분석을 위한 양적 연구에 LDA 기반의 알고리즘인 Netminer<sup>9)</sup>를 이용한 여러 연구들(박영훈·김장식·송향근, 2019; 박영훈, 2020; 정연희, 2020; 방성원, 2021; 김동아, 2022; 조남현·장경완, 2023)이 발표되었다. 본 연구 역시 연구 주제들을 도출하기 위하여 Netminer를 사용하였다. 다음으로 시간의 흐름에 따른 연구 주제들의 추이 변화를 분석하기 통계 분석 소프트웨어인 ‘StataSE 18’을 이용하여 시계열 회귀분석을 하였다.

### 3. 연구 결과

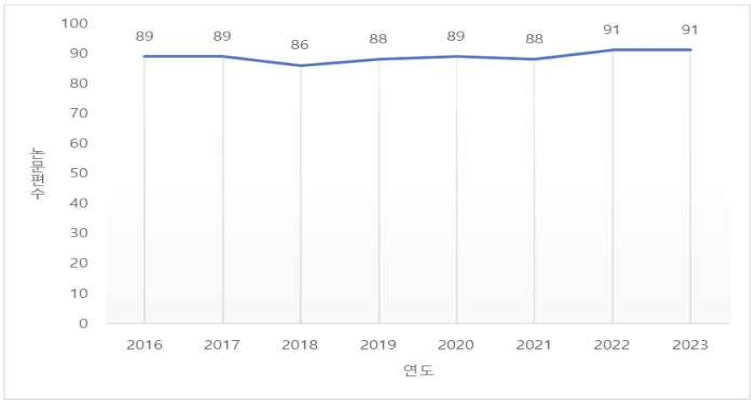
#### 3.1 논문 수 추이

##### 3.1.1 연도별 추이

2016년에서 2023년까지 연도별 논문 추이는 전체 논문의 수는 전반적으로 큰 변동 없이 비교적 일정했다. 조금 더 자세히 살펴보면 2018년 이후부터 서서히 증가하는 경향을 보여주고 있으며 최근 몇 년 동안에는 다소 증가하는 추세를 보였다.

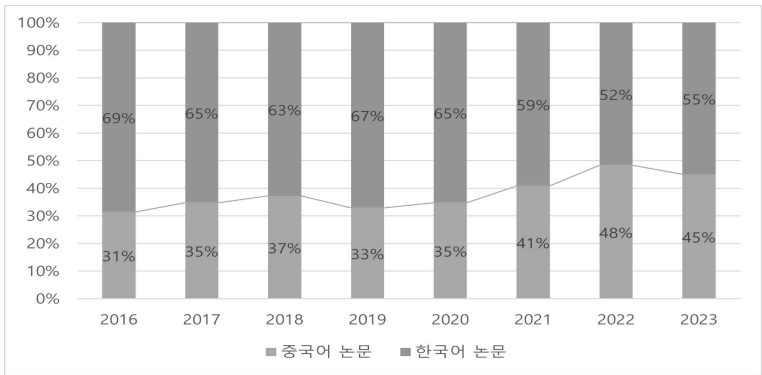
---

9) ㈜사이람에서 개발한 네트워크 & 소셜 빅데이터 분석 전문 소프트웨어로 한국어 분석이 가능해 자료 수집과 분석에 장점이 있다.



<그림 2> 2016~2023년 연도별 논문 수

그러나 언어별 비율의 변화는 차이가 있었다. 2016년부터 2023년까지 중국어와 한국어로 작성된 논문의 비율을 살펴보면, 두 언어의 사용 비율에 뚜렷한 변화가 나타난다. <그림 3>은 해당 기간 중국어와 한국어 논문의 비율 변화를 보여준다.



<그림 3> 2016~2023년 중한 논문 퍼센트

중국어 논문의 비율은 전반적으로 꾸준히 증가하는 경향을 보인다. 특히 2020년 이후로 그 비율이 급격히 상승하여, 2022년에 가장 많은 논문 비율을 기록했다. 2023년에는 45%로 소폭 감소했지만, 여전히

높은 수준을 유지하고 있다. 반면, 한국어 논문의 비율은 감소하는 추세를 보인다. 특히 2020년 이후로 감소 폭이 커졌으며, 2021년에는 그 비율이 52%로 가장 낮았다. 2023년에는 55%로 다소 증가했지만, 초기의 높은 비율을 회복하지 못한 상태이다.

이러한 변화는 학술지 논문 저자들의 언어 선택의 변화, 연구 주제의 전환, 학술지의 언어 정책 등의 여러 요인에 의해서 영향을 받을 수 있다. 앞으로도 이러한 언어 사용 패턴의 변화를 계속해서 관찰하고 분석할 필요가 있다.

### 3.1.2 연도별 회귀분석

<표 2>는 2016년부터 2023년까지 중국어 논문과 한국어 논문에 대한 회귀분석 결과이다. 분석은 전체 기간과 코로나 발생 이전과 이후로 나누어 이루어졌다.

<표 2> 시계열 분석

	$\beta$	$t$	DW	$p$
<b>2016-2023</b>				
중국어 논문	0.0209**	3.96	1.648	0.007
한국어 논문	-0.0209**	-3.96	1.648	0.007
<b>2016-2020</b>				
중국어 논문	0.0049	0.66	2.197	0.559
한국어 논문	-0.0049	-0.66	2.197	0.559
<b>2021-2023</b>				
중국어 논문	0.0207	0.67	3	0.625
한국어 논문	-0.0207	-0.67	3	0.625

\*\*\*  $p < .001$ , \*\*  $p < .05$ , \*  $p < .1$

전체 기간(2016~2023년) 동안 중국어 논문 비율은 통계적으로 유의미하게 증가한 것으로 나타났다( $\beta = 0.0209$ ,  $p = 0.007$ ). 반면 한국어 논문의 비율은 통계적으로 유의미하게 감소했다( $\beta = -0.0209$ ,  $p = 0.007$ ). 더빈-왓슨 값<sup>10)</sup>(DW = 1.648)은 잔차 간의 경미한 자기상관

10) 더빈-왓슨(Durbin-Watson, DW) 값은 회귀분석에서 잔차(오차 항) 간의 자기

을 나타내며, 모델의 독립성에는 큰 문제가 없음을 보여준다. 따라서 앞으로도 해당 학술지에서는 한국어보다는 중국어로 쓴 논문이 발표될 가능성이 높다는 것을 제시한다.

코로나 이전 기간(2016~2020년)의 경우 중국어 논문과 한국어 논문의 비율 변화는 통계적으로 유의미하지 않았다(각각  $p = 0.559$ ). 따라서 이 시기에는 두 언어별 논문의 비율 변화는 거의 없었던 것으로 보인다. DW 값이 2.197로 나타나 잔차 간의 자기상관이 거의 없었다. 코로나 이후 기간(2021~2023년)의 경우 역시 중국어 논문과 한국어 논문의 비율 변화는 통계적으로 유의미하지 않았다(각각  $p = 0.625$ ). 비율의 변화가 있었지만, 그 변화가 통계적으로 신뢰할 만한 수준은 아니었으며 DW 값이 3으로 높게 나타나서 잔차 간의 높은 자기상관 가능성이 보인다. 그러므로 전체 기간에 대한 회귀분석 결과는 유의미하지만, 하위 기간별 분석에서는 통계적으로 유의미한 차이가 나타나지 않았다. 연구 해석의 정확성을 높이기 위해서 추가적인 데이터의 검토가 필요해 보인다.

## 3.2 언어 네트워크 분석 결과

### 3.2.1 키워드 출현 빈도분석

분석 대상인 711편의 연구 논문의 제목에 나타난 명사는 모두 1,965개였으며 이것을 다시 전처리 과정을 통해 정제하였다. 전처리 과정을 거친 데이터를 대상으로 빈도분석을 시행하여 추출된 단어의 수는 1,761개이다. 상위 빈도 30위 이내의 고빈도 단어는 <표 3>과 같다. ‘번역’, ‘교육’, ‘교수’, ‘문화’, ‘문학’ 등이 최상위 빈도로 나타났고 이 단어들은 잠재적 연구 주제가 될 가능성이 높다. 그러나 문서상에서 특정

---

상관성을 측정하는 통계 지표이다. 이 값은 주로 선형 회귀 모델에서 잔차가 독립적인지를 검정하는 데 사용이 된다. DW 값은 2에 가까울수록 잔차가 독립적이라는 것을 나타내며, 이는 회귀 모델의 기본 가정이 충족되고 있음을 의미한다. 따라서 모델의 예측이 더 신뢰할 수 있으며, 독립성 가정이 충족되어 미래 예측에 유리하다(Durbin, J., & Watson, G. S., 1992).

단어가 많이 나타난다고 해도 핵심어가 아닐 수도 있음에 주의해야 한다. 즉 문서에는 자주 출현하지만, 주제로 보기에 어려운 보편적인 단어일 수도 있으므로 이어지는 토픽 모델링과 언어 네트워크 분석을 통해서 연구 주제와 핵심어를 재차 확인하는 과정이 필요하다.

고빈도 단어 분석 결과 번역(224회), 교육(172회), 교수(167회), 문화(145회), 문학(87회) 등이 가장 많이 사용된 단어로 나타났다. 이러한 단어들은 논문 제목에서 자주 등장하며, 이들이 연구의 주요 주제나 관심사를 반영하고 있음을 시사한다. 중간 빈도의 단어로는 대학교(49회), 대조(45회), 여성(44회), 특징(44회), 오류(42회) 등이 있었다. 이들 어휘는 특정 세부 주제나 연구 초점에 대한 관심을 나타낼 수 있다. 교수법(31회), 이론(31회), 중국인(31회) 등의 단어들은 상대적으로 낮은 빈도를 보이지만, 연구 제목에서 나타날 만큼 중요성이 있다고 할 수 있다.

번역이 가장 높은 빈도를 차지하고 있다는 점에서, 번역 관련 연구가 매우 활발하게 이루어지고 있음을 알 수 있다. 또한 고빈도 단어 목록에서 다양한 분야와 관련된 용어들(의미, 소셜, 여성, 기능, 오류, 전략 등)이 나타나는 것은 연구 주제가 광범위하며 다양한 학문적 논의가 이루어지고 있음을 보여준다. 마찬가지로 교수나 교육과 같은 단어가 상위권에 위치하는 것도 교육 관련 연구, 특히 교육 방법론이나 교수법에 관한 관심이 높음을 보여준다.

<표 3> 상위 30위 고빈도 단어

순위	단어	Frequency	순위	단어	Frequency
1	번역	224	16	대조	45
2	교육	172	17	여성	44
3	교수	167	18	특징	44
4	문화	145	19	오류	42
5	문학	87	20	전략	41
6	비교	87	21	기능	40
7	의미	79	22	어휘	39
8	소설	75	23	한중	39
9	중국	74	24	이미지	38
10	과정	59	25	의식	34
11	학습자	58	26	전공	34
12	중국어	57	27	학습	34
13	방식	52	28	교수법	31
14	언어	52	29	이론	31
15	대학	49	30	중국인	31

<표 3>에서 제시한 고빈도 단어들을 직관적이고 인지적으로 파악하기 위해 워드 클라우드를 통해서 시각화하였다. <그림 4>는 상위 250개 키워드로 구성된 워드 클라우드이다. 이를 바탕으로 '번역'과 '교육'이 이 텍스트의 주요 주제임을 알 수 있었다. 특히 번역이 중심에 자리하고 있어, 해당 자료가 번역 또는 교육 자료에 관한 연구일 가능성이 크다. 또한 '문화', '비교', '의미' 등의 단어들도 나타나는 것으로 보아서 다양한 학문적 주제를 다루고 있다는 것을 알 수 있다.



<그림 4> 워드 클라우드

### 3.2.2 중심성 분석 결과

텍스트 내의 주제와 구조를 보다 체계적으로 이해하고 중요한 패턴을 식별하기 위해서 언어 네트워크 분석을 하였다. 세부적으로는 우선 각 토픽에서 중요한 위치를 차지하고 있는 키워드 간의 연결 관계를 더 정밀하게 분석하기 위해서 동시 출현 단어 분석을 수행하고 핵심적인 개념이나 주제를 명확히 하기 위해서 중심성을 분석하였고, 그 결과를 키워드 네트워크 수형도로 도식화하여 의미 관계를 살펴보았다.

#### (1) 동시 출현 단어 분석

동시 출현 단어 분석은 텍스트 내에서 특정 단어들이 함께 나타나는 빈도를 측정하여, 단어 간의 관계를 시각화하는 방법이다. 이는 텍스트를 분석하기 위해 텍스트에 동시에 나타나는 단어 쌍의 패턴을 활용하는 방법으로 단어 쌍 간의 연결 강도를 통해 유사도를 산출하고 이를 기반으로 주제를 묶거나 네트워크 맵(map)으로 표현하는 분석 기법이다(허영수, 2023:651). 또한 특정 분야 문헌들의 유형 및 경향을 살펴보는 데에 적합한 연구 방법으로 알려져 있다(김은실·강승혜, 2019:27).

본 연구에서 분석한 상위 30위 동시 출현 단어쌍은 <표 4>와 같다. <표 4>는 논문 제목과 초록에서 자주 함께 나타난 상위 30위 동시 출현 단어쌍이다. 이 표에서 ‘Source’는 첫 번째 단어, ‘Target’은 두 번째 단어를 의미하며, ‘강도(weight)’는 각 단어쌍이 얼마나 자주 함께 나타났는지를 수치로 표현한 것이다. 값이 클수록 두 단어가 자주 같이 사용되었음을 의미한다. 가장 자주 함께 사용된 단어쌍은 ‘문화-교육’(강도 18), ‘중국인-학습자’(강도 18), ‘교수-방식’(강도 17), ‘번역-전략’(강도 16) 등이다. 이는 연구에서 문화와 교육이 밀접하게 관련된 주제들로 다루어졌으며, 연구 대상이 중국인 학습자였음을 보여준다. 또한, ‘인재-양성’(강도 16)과 ‘플립-학습’(강도 14)과 같은 단어쌍은 연구자들이 학문 목적 학습자들과 최신 교육 방법에 대한 관심이 높았음을 나타낸다. 중간 빈도의 단어쌍으로는 ‘번역-과정’(강도 8), ‘교수-개혁’(강도 8) 등이 있으며, 이는 특정 연구 주제에서 중

요한 요소로 다루어졌음을 시사한다. 이러한 단어쌍들은 각각의 연구 분야나 주제에서 중요한 키워드로 사용된 것으로 보이며, 해당 주제에 대한 학문적 관심을 반영한다.

<표 4> 상위 30위 동시 출현 단어쌍

순위	공출현 단어쌍		강도 (weight)	순위	공출현 단어쌍		강도 (weight)
	Source	Target			Source	Target	
1	문화	교육	18	16	번역	과정	8
2	중국인	학습자	18	17	번역	소개	8
3	교수	방식	17	18	교수	개혁	8
4	번역	전략	16	19	초급	학습자	7
5	인재	양성	16	20	과정	개혁	7
6	중국	문화	14	21	사용	실태	7
7	플립	학습	14	22	과정	교수	7
8	병렬	말뭉치	10	23	접속	어미	7
9	교육	내용	10	24	번역	교육	7
10	문화	교류	10	25	발음	교육	7
11	번역	교수	9	26	여성	문학	6
12	화용	기능	9	27	교수	내용	6
13	한중	번역	9	28	작가	의식	6
14	의미	기능	8	29	번역	이론	6
15	중국	대학	8	30	번역	오류	6

(2) 중심성 분석

연구 주제를 정확하게 반영할 수 있는 핵심어 추출을 위해서 중심성 분석을 하였다. 중심성 분석을 통해서 전체 네트워크상에서 중요한 위치나 영향을 주는 단어를 찾는데, 이를 측정하는 지표로 연결 중심성(Degree centrality), 매개 중심성(Betweenness centrality), 근접 중심성(Closeness centrality)을 구하고 비교하였다.

<표 5> 중심성 상위 20개 키워드

순위	연결중심성 (Degree Centrality)		매개중심성 (Betweenness Centrality)		근접중심성 (Closeness Centrality)	
	키워드	값	키워드	값	키워드	값
1	번역	0.041219	문화	0.086617	비교	0.245704
2	문화	0.033453	번역	0.086022	번역	0.244982
3	교수	0.032855	교수	0.069492	교수	0.244862
4	비교	0.028674	문학	0.062701	문화	0.238214
5	소셜	0.026882	교육	0.052998	교육	0.233597
6	교육	0.026284	의미	0.052108	의미	0.231971
7	의미	0.026284	중국	0.042573	언어	0.219789
8	문학	0.018519	비교	0.042114	활용	0.219741
9	언어	0.016129	소셜	0.040301	과정	0.219163
10	여성	0.014337	언어	0.031145	문학	0.215713
11	특징	0.014337	중국어	0.02756	특징	0.215573
12	관점	0.01374	여성	0.024176	교수법	0.215341
13	대조	0.01374	방식	0.021828	오류	0.212144
14	중국	0.01374	어휘	0.020446	수업	0.211695
15	활용	0.01374	특징	0.020069	방식	0.211651
16	오류	0.012545	활용	0.0181	관련	0.211516
17	방식	0.011947	한중	0.01723	이론	0.211293
18	의식	0.01135	학습자	0.017152	전략	0.209391
19	중국어	0.01135	의식	0.015386	중국어	0.20891
20	표현	0.01135	구조	0.015189	특성	0.208474

연결 중심성(Degree Centrality)은 각 단어가 다른 단어들과 얼마나 많이 직접적으로 연결되어 있는지를 나타낸다. 값이 크면 해당 단어가 다른 단어와 직접적으로 연결되어 있어 네트워크에서 중요한 위치에 있음을 의미한다. 번역(0.041219), 문화(0.033453), 교수(0.032855) 등이 높은 연결 중심성을 나타내며, 이는 이 단어들이 다른 단어들과 자주 연결되어 논문 제목이나 초록에서 중요하게 다뤄졌음을 의미한다. 매개 중심성(Betweenness Centrality)은 특정 단어가 네트워크 내에서 다른 단어들 사이에 연결을 얼마나 중개하는 역할을 하는지를 나타낸다. 값이 클수록 해당 단어가 네트워크 내에서 중요한 연결 지점을 형성하고 있음을 의미한다. 문화(0.086617), 번역(0.086022), 교수(0.069492) 등이 높은 매개 중심성을 보이며, 이는 이러한 단어들이 다른 단어와 단어 사이의 연결 과정에서 중요한 중재자의 기능을 하고 있다. 근접 중심성(Closeness Centrality)은 특정 단어가 네트워크 내 다른 단어들과 얼마나 가까운지(즉, 얼마나 빠르게 다른 단어들과 연결될 수 있는지)를 측정한다. 값이 클수록 해당 단어가 네트워크의 중심

에 위치하고 있음을 의미하는데 비교(0.245704), 번역(0.244892), 교수(0.244862) 등이 높은 근접 중심성을 보였다. 이는 이러한 단어들이 네트워크의 중심에 위치하며, 다른 단어들로부터 쉽게 접근이 가능한 위치에 있음을 알 수 있다.

번역은 세 가지 중심성 지표에서 모두 높은 값을 보이며, 이는 논문 제목과 초록 내에서 매우 중요한 핵심어로 사용되고 있음을 의미한다. 문화, 교육, 비교도 모든 지표에서 높은 점수를 기록하여 중요한 역할을 하고 있다. 문화는 매개 중심성에서 가장 높은 점수를 기록하여, 다른 주제 간 연결에서 중요한 역할을 하고 있음을 보여준다. 높은 중심성을 가진 단어들은 연구에서 더 많은 주목을 받고 있거나, 여러 연구에서 중요하게 다뤄진 주제일 가능성이 크다. 따라서 이러한 단어들을 중심으로 추가적인 연구를 계획하거나, 기존 연구의 주제 사이의 관계를 재평가할 수 있다.

중심성 분석을 바탕으로 키워드 네트워크를 시각화하였다. <그림 5>에서 붉은 점은 핵심 단어 즉 노드(node)를 의미한다. 예를 들어 ‘번역’, ‘문화’, ‘교육’은 핵심 키워드로 하나가 각각의 노드를 표시한다. 그리고 이러한 단어들 사이의 연결을 나타내는 것이 엣지(Edge)이다. 엣지가 많은 노드는 다른 키워드와 자주 연결되며, 전체 의미 네트워크에서 중요한 역할을 하는 키워드를 나타낸다.



또 다양한 키워드들이 연결된 모습을 통해 연구 주제의 폭과 깊이를 확인할 수 있다. 예를 들어 ‘비교’, ‘기능’, ‘구조’와 같은 키워드들은 학문적 접근법을 반영한 것으로 보인다.

### 3.3. 토픽 모델링 분석 결과

#### 3.3.1. 토픽 모델링 분석 결과

토픽 모델링을 수행하기 전에 최적의 토픽 개수(Number of Topics)를 찾기 위해서  $c_v$ (Coherence Value)<sup>11)</sup> 값을 구했다. 토픽 개수는 첫 번째 열에 나타난 것처럼 4개부터 25개까지를 설정하였다.  $\alpha$  (Alpha)는 주제 분포의 과정을 제어하는 매개변수로, 각 문서에 할당되는 주제의 분포를 나타낸다.  $\alpha$  값이 클수록 문서에 더 많은 주제가 골고루 분포되며, 값이 작을수록 한두 개의 주제가 집중되는 경향이 있다. 이에 선행연구(최진수·정혜원, 2022:14)와 문서의 양을 고려하여  $\alpha$  값을 0.1로 설정했는데 이는 중간 정도의 주제 분포를 가정한다.  $\beta$  (Beta)는 단어 분포의 과정을 제어하는 매개변수로, 각 주제 내에서 단어들어 어떻게 분포되는지를 나타낸다.  $\beta$  값이 작을수록 각 주제는 소수의 단어에 집중되고, 값이 클수록 단어들어 골고루 분포된다.  $\alpha$  값과 동일한 기준을 따라 0.001로 설정하였다. 반복 횟수(Iterations)는 각 주제 모델링 시뮬레이션에서 실행된 반복 횟수를 나타낸다. 표에서는 모든 시뮬레이션에서 1,000회의 반복을 수행했다. 이 반복 횟수는 모델의 수렴을 돕고, 안정적인 결과를 얻기 위해 설정된 것이다. 그리고 <표 6>에서 주제 개수가 9일 때  $c_v$  값이 0.61로 가장 높게 나타나 주제의 일관성을 보인다. 이를 참조하여 주제의 개수를 9개로 지정하고 이를 바탕으로 핵심 키워드를 확인하였다.

11) 특정 주제와 관련이 있는 단어들의 공출현 정도를 평가하여 해당 주제가 의미론적으로 얼마나 일관성이 있는지 판단하는 지표이다(Röder, M., Both, A., & Hinneburg, A., 2015). 값이 클수록 해당 주제가 더 일관성 있고 언어적으로 의미가 있다는 것을 나타낸다.

<표 6> 토픽 개수별 c\_v값

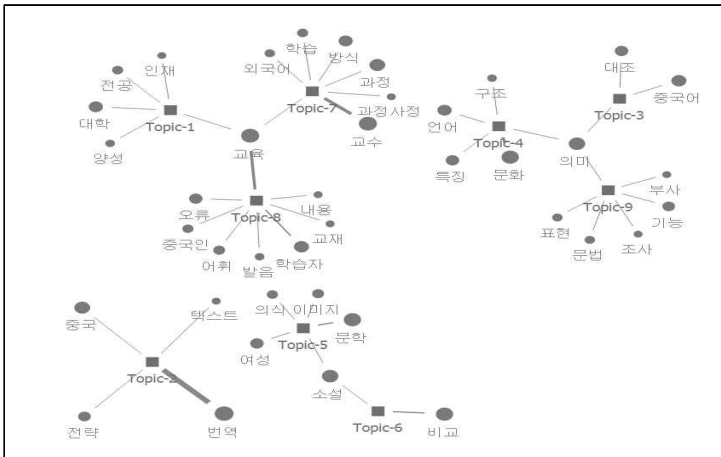
토픽 개수	$\alpha$	$\beta$	반복 횟수	c_v
4	0.1	0.01	1000	0.437
5	0.1	0.01	1000	0.542
6	0.1	0.01	1000	0.583
7	0.1	0.01	1000	0.468
8	0.1	0.01	1000	0.502
9	0.1	0.01	1000	0.61
10	0.1	0.01	1000	0.532
11	0.1	0.01	1000	0.497
12	0.1	0.01	1000	0.533
13	0.1	0.01	1000	0.496
14	0.1	0.01	1000	0.53
15	0.1	0.01	1000	0.513
16	0.1	0.01	1000	0.542
17	0.1	0.01	1000	0.545
18	0.1	0.01	1000	0.55
19	0.1	0.01	1000	0.46
20	0.1	0.01	1000	0.529
21	0.1	0.01	1000	0.53
22	0.1	0.01	1000	0.522
23	0.1	0.01	1000	0.536
24	0.1	0.01	1000	0.56
25	0.1	0.01	1000	0.522

<표 7>에서 확인할 수 있듯이, 연구는 다양한 주제를 포괄하고 있다. 전공/대학, 번역, 대조, 비교, 문학, 교수, 학습자 교육, 문법 등 다양한 주제가 다루어졌으며, 이는 연구의 폭넓은 범위를 보여준다.

<표 7> 토픽별 키워드, 연구 주제, 구성 비율

Topic	1 위 키워드	2 위 키워드	3 위 키워드	4 위 키워드	5 위 키워드	연구 주제	구성 비율(%)
Topic 1	교육 (0.043)	전공 (0.042)	양성 (0.036)	대학 (0.035)	인재 (0.028)	전공/대학	9.0%
Topic 2	번역 (0.234)	중국 (0.031)	전략 (0.028)	텍스트 (0.025)	문화 (0.018)	번역	12.5%
Topic 3	대조 (0.052)	중국어 (0.048)	의미 (0.047)	함종 (0.023)	한자어 (0.02)	대조	11.4%
Topic 4	문화 (0.124)	특징 (0.047)	언어 (0.032)	의미 (0.028)	구조 (0.027)	문화	7.6%
Topic 5	문학 (0.056)	소설 (0.039)	여성 (0.043)	이미지 (0.035)	의식 (0.033)	문학	16.2%
Topic 6	비교 (0.059)	소설 (0.039)	실화 (0.019)	시대 (0.019)	외래어 (0.015)	비교	8.6%
Topic 7	교수 (0.147)	과정 (0.045)	학습 (0.033)	방식 (0.03)	교육 (0.028)	교수	13.1%
Topic 8	교육 (0.123)	학습자 (0.067)	요류 (0.047)	어휘 (0.038)	중국인 (0.036)	학습자 교육	10.5%
Topic 9	기능 (0.052)	부사 (0.03)	조사 (0.029)	문법 (0.027)	의미 (0.027)	문법	11.1%

토픽 모델링 결과를 시각화하면 <그림 6>과 같다. 각 토픽은 파란색 사각형 노드로 표시되어 있으며, 각 토픽에 해당하는 번호가 기재되어 있다(예: Topic-1, Topic-2 등). 각 토픽은 분석을 통해 도출된 특정 주제를 나타내며, 관련된 키워드들로 둘러싸여 있다. 각 키워드는 붉은 점으로 표시되어 있고 키워드들의 크기와 위치는 해당 키워드의 중요도를 나타내며, 크기가 클수록 그 토픽 내에서 중요도가 높은 키워드를 의미한다. 토픽과 키워드 간의 선(엣지)은 해당 키워드가 토픽과 얼마나 밀접하게 연관되어 있는지를 나타낸다. 선의 두께가 두꺼울수록 해당 키워드와 토픽 간의 연결이 강하고, 자주 함께 나타났음을 의미한다. 예를 들어, Topic-1에서는 교육, 전공, 양성, 대학, 인재가 주요 키워드로 나타난다.



<그림 6> 토픽 모델링 시각화

분석 결과 총 4개의 군집이 있으며, 각각의 군집은 특정 주제를 중심으로 관련 단어들이 연결된 형태로 나타나 있다는 것을 알 수 있다. Topic-1, 7, 8이 연결되어 있는 군집은 ‘대학’, ‘과정’, ‘교수’, ‘학습자’ 등의 단어와 연결되어 있다. 이 군집은 주로 ‘교육’ 관련 주제를 다루고 있는 것으로 보이며, 인재 양성, 한국어 교육, 교육과정 등과 관련된 내용을 포함하고 있다. Topic-3, 4, 9가 연결된 군집은 ‘의미’, ‘문화’, ‘기

능’과 같은 단어들이 중심에 있다. 이 군집은 ‘언어’와 관련이 있는 주제를 다루고 있는 것으로 보인다. 언어의 의미론적 측면이나 문화적 문맥, 문법 등의 주제와 관련된 연구들이 연결되어 있었다. Topic-5, 6은 ‘여성’, ‘문학’, ‘비교’와 같은 핵심어들이 포함되어 있다. 이 군집은 ‘문학’과 관련된 주제로 보인다. 특정한 문학 텍스트와 문학 작품의 비교 등과 관련된 연구들이다. Topic-2는 독립적인 주제로 ‘번역’, ‘전략’, ‘중국’와 같은 단어들이 연결되어 있다. 이 4개의 군집의 비율을 살펴보면 ‘교육’과 관련된 군집이 32%로 가장 많고 다음으로 ‘언어’와 관련된 군집의 비율이 30.1%, 그리고 ‘문학’ 24.4%, ‘번역’ 12.5% 순으로 나타났다.

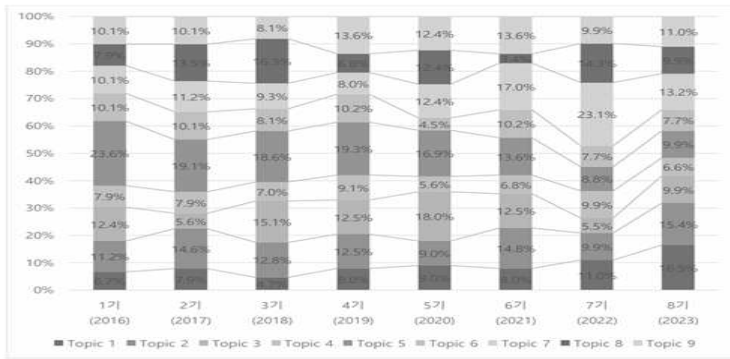
### 3.3.2 연구 주제의 변화 양상 분석

다음으로 시간의 변화에 따라 연구 주제가 양적으로 어떻게 변화하였는지를 살펴보았다. <표 8>은 관련 논문들을 주제와 발표 연도를 기준으로 분류한 것이다.

<표 8> 토픽별/시기별 논문 발행 편수

Topic	1기 (2016)	2기 (2017)	3기 (2018)	4기 (2019)	5기 (2020)	6기 (2021)	7기 (2022)	8기 (2023)	합계
Topic 1 전공/대학	6	7	4	7	8	7	10	15	64
Topic 2 번역	10	13	11	11	8	13	9	14	89
Topic 3 대조	11	5	13	11	16	11	5	9	81
Topic 4 문화	7	7	6	8	5	6	9	6	54
Topic 5 문학	21	17	16	17	15	12	8	9	115
Topic 6 비교	9	9	7	9	4	9	7	7	61
Topic 7 교수	9	10	8	7	11	15	21	12	93
Topic 8 학습자 교육	7	12	14	6	11	3	13	9	75
Topic 9 문법	9	9	7	12	11	12	9	10	79
합계	89	89	86	88	89	88	91	91	711

그리고 분석의 대상이 되는 2016년부터 2023년까지 연도별로 연구 주제의 구성 비율을 구하고 이를 다시 <그림 7>과 같이 시각화였다.



<그림 7> 시기별 연구 주제의 변화

시기별 연구 주제 변화가 심한 주제는 다음과 같다. Topic-5 ‘문학’은 2016년(1기)에는 전체 연구의 23.6%로 가장 높은 비율을 차지했으나, 2019년(4기)에는 19.3%로 감소하고, 이후 2022년(7기)에는 8.8%로 가장 낮은 비율을 기록했다. 2023년(8기)에는 9.9%로 다소 회복했지만, 초기의 높은 비율과 비교하면 상당한 감소세를 보인다. Topic-6 ‘비교’는 2016년(1기)부터 2019년(4기)까지 10% 전후로 안정된 비율을 보이다가 2020년(5기)에는 4.5%로 급감했다. 이후 2021년(6기)에 다시 과거와 비슷한 10.2%로 증가했다가 2022년 이후에는 7.7%의 다소 감소한 비율을 나타낸다. 주목할 점은 2020년(5기) 때의 감소이다. 다른 시기에는 안정세를 보였는데 유독 특정한 시기에 감소세를 보인 것은 연구 외적 요인에 영향을 받았을 가능성이 있다. 이 시기에 코로나가 발생했다는 점이 비교 연구에 영향을 주었을 가능성이 있다. Topic-7 ‘교수’는 2016년(1기)에 10.1%로 시작하여 점차 증가하는 모습을 보이다가 2022년(7기)에는 23.1%로 정점을 찍었다. 하지만 2023년(8기)에는 13.2%로 다시 감소하는 등 변화의 폭이 큰 편이다. Topic-8 ‘학습자 교육’도 2016년(1기)에는 7.9%로 시작하여 2018년(3기)에는 16.3%로 최고치를 기록한 후에는 감소와 증가를 반복하는 추세를 보인다.

이러한 변화의 원인은 연구 경향 및 사회적 관심의 변화가 있다는 것을 의미한다. ‘Topic 5 문학’의 변화는 문학 연구가 초기에는 큰 관심

을 받았으나, 최근 들어 다른 학문적 관심사인 교육, 번역, 한국어 교육 등에 밀려 상대적으로 중요도가 낮아졌기 때문일 수 있다. 이는 사회적 이슈나 새로운 연구 주제의 부각으로 인해 연구자들의 관심이 다른 분야로 이동한 결과이거나 해당 분야의 전공자들의 감소 등에서 원인을 찾을 수 있을 것이다. Topic 7 ‘교수’와 Topic 8 ‘학습자 교육’의 경우, 교육 정책의 변화나 국제적 환경 변화에 따라 학습자 연구가 집중되었을 가능성이 크다. 예를 들어, 한국어 교육은 최근 몇 년 동안 국제적으로 한국 문화와 한류의 영향력 확대에 의해 중요성이 증가했으며, 이에 대한 연구가 폭발적으로 증가했다가, 일정 기간 이후 어느 정도 안정된 수요에 따라 감소했을 수 있다.

다음 단계로 시간의 변화 속에서 연구자들이 어떤 주제에 관심이 많았는지 살펴보기 위해서 시계열 회귀분석을 하였다. 이 분석을 통해서 선행연구(Griffiths, T. L., & Steyvers, M, 2004; 김태경·최희련·이홍철, 2016; 박종순·김창식, 2019) 등에서 구분한 인기 주제(Hot Topic), 비인기 주제(Cold Topic), 중립 주제(Neutral Topic)를 확인해 보았다.

<표 9> 연구 주제별 시계열 분석

Topic		$\beta$	t	DW	p
Topic 1	전공/대학	0.0113**	3.1	1.51	0.021
Topic 2	번역	0.001	0.26	2.97	0.807
Topic 3	대조	-0.0025	-0.34	2.1	0.747
Topic 4	문화	-0.00038	-0.16	2.88	0.875
Topic 5	문학	-0.1962***	-7.23	1.89	0.000
Topic 6	비교	-0.0034	-1.1	3.35	0.314
Topic 7	교수	0.0129*	2.07	1.89	0.084
Topic 8	학습자 교육	-0.00179	-0.25	2.92	0.811
Topic 9	문법	0.00245	0.79	2.16	0.458

\*\*\*p<.001, \*\*p<.05, \*p<.1

회귀 계수( $\beta$ )는 각 주제의 시간에 따른 변화 추세를 나타낸다.  $\beta$  값이 양수일 경우, 해당 주제의 비율이 시간이 지남에 따라 증가하고 있

음을 의미하며, 음수일 경우 감소하고 있음을 의미한다. 예를 들어, Topic-1 전공/대학의  $\beta$  값은 0.0113으로 양수이며, 이는 이 주제가 시간이 지남에 따라 증가하는 추세임을 나타낸다. 반면, Topic-5 문학의  $\beta$  값은 -0.1962로 음수이며, 이는 이 주제가 시간이 지남에 따라 감소하는 추세를 보이고 있음을 시사한다.  $t$  값은  $\beta$  값의 통계적 유의성을 테스트하는 지표이다.  $t$  값이 클수록 해당 회귀 계수가 통계적으로 유의미하다는 것을 나타낸다.  $p$  값은 회귀 계수가 통계적으로 유의미한지를 판단하는 데 사용된다. 본 연구에서는  $p$  값이 0.1, 0.05, 0.001 이하이면 통계적으로 유의미하다고 간주한다.

Topic-1 전공/대학( $\beta = 0.0113$ ,  $p = 0.021$ )은 시간이 지남에 따라 이 주제의 비율이 증가했다. 통계적으로 유의미한 추세를 보이고 자기상관성도 미미하게 나타났다. 이는 전공/대학 관련 연구가 최근 들어 점점 더 중요한 주제로 부각되고 있음을 시사한다. Topic-7 교수( $\beta = 0.0129$ ,  $p = 0.084$ )는  $p$  값이 0.084로 0.1보다 작으며 통계적으로 유의미한 증가 경향을 확인할 수 있었다. 반면 Topic-5 문학( $\beta = -0.1962$ ,  $p = 0.000$ )은  $p$  값이 0.001보다 작아 시간이 지남에 따라 유의미하게 감소하는 추세를 보인다. 이는 문학 관련 연구가 최근에는 상대적으로 중요도가 낮아졌음을 시사한다.

대부분의 주제에서 DW 값이 2에 가까운 값을 보이며, 이는 잔차의 자기상관이 크지 않음을 의미한다. 다만, ‘번역’, ‘비교’, ‘한국어 교육’ 등의 주제의 DW 값이 다소 높은 편이어서, 잔차 간에 양의 자기상관이 있을 가능성이 있다. 정리하면 ‘전공/대학’, ‘교수’ 등의 주제가 통계적으로 유의미한 증가 추세를 보이는 인기 주제이며 ‘문학’이 감소세에 있는 비인기 주제임을 알 수 있다. 나머지 주제들은 증감이 통계적으로 유의하게 나타나지 않아 중립 주제라고 할 수 있다.

#### 4. 결론 및 제언

본 연구에서는 2016년부터 2023년까지 중국 학술지 ‘한국어 교수와 연구’에 실린 논문의 언어 사용 및 주제 변화를 분석하였다. 연구 결과,

전체 논문의 수는 비교적 일정하게 유지되었으나, 동일 기간에 논문에 사용된 언어의 변화가 두드러졌다. 중국어 논문의 비율은 지속적인 증가세를 보였는데 특히 2020년 이후 큰 상승이 나타났다가 2023년에는 소폭 감소했다. 반면, 한국어 논문의 비율은 지속적인 감소 추세를 보였다.

동일한 자료를 대상으로 회귀분석을 한 결과, 중국어 논문의 비율은 통계적으로 유의미하게 증가했으나 한국어 논문의 비율은 유의미하게 감소했다. 하지만, 코로나 이전과 이후로 나누어 분석한 결과, 하위 기간별로는 통계적으로 유의미한 차이가 없었다. 이는 전반적인 언어 사용 변화는 명확했지만, 유행병과 같은 외생 변수의 영향이 생각보다 크지 않았음을 보여준다. 세부 시기의 변화는 다른 학술지 등을 포함한 추가 연구를 통해서 살펴볼 수 있을 것이다.

다음으로 여러 주제를 구성하는 핵심 키워드를 확인하기 위해서 키워드 분석을 하였고 ‘번역’, ‘교육’, ‘교수’, ‘문화’, ‘문학’ 등이 주요 연구 주제임이 확인되었다. 또한 키워드 사이의 관계를 파악하기 위해서 언어 네트워크 분석을 하였다. 먼저 동시 출현 단어를 분석하고 다음으로 중심성을 살펴보았다. 분석 결과, ‘번역’, ‘교육’, ‘문화’가 핵심적인 연구 주제임이 드러났으며, 이들 간의 관계가 강하게 연결되어 있음을 확인할 수 있었다.

앞선 언어 네트워크 분석을 기초로 연구 주제를 찾기 위해서 토픽 모델링 분석을 하였고 총 9개의 토픽(‘전공/대학’, ‘번역’, ‘대조’, ‘문화’, ‘문학’, ‘비교’, ‘교수’, ‘학습자 교육’, ‘문법’)이 도출되었다. 이것은 해당 학술지가 다양한 연구 주제를 다루고 있음을 보여준다. 다음으로 해당 토픽과 관련된 키워드를 시각화하였다. 분석 결과 총 4개의 군집을 확인하였는데 각각의 군집이 ‘교육’, ‘언어’, ‘문학’, ‘번역’과 관련된 주제라는 것을 확인하였다.

또한 연구 주제의 변화 양상을 분석한 결과, 특정 주제의 비율이 시기별로 변화가 있었다. ‘문학’은 초기에는 높은 비율을 차지했으나 지속적인 감소세를 보였고 2022년에 가장 낮게 나타났다. 이러한 추세는 통계적으로도 증명되었다. 시계열 회귀분석 결과, ‘문학’의 감소 추세는 통계적으로 유의미하였는데 이는 해당 학술지의 연구자들이 문학

관련 연구가 활발하지 않음을 의미한다. 반대로 ‘전공/대학’과 ‘교수’와 관련된 연구는 증가세를 보였다. 이는 학문 목적 학습자를 대상으로 한 연구의 증가를 시사한다. 통계적 분석 결과 역시 유의미해서 이와 관련된 연구가 증가 추세를 보이는 인기 주제임을 알 수 있었다. 다른 연구 주제들은 변동이 크지 않아 일정한 비율을 유지하거나 증감을 반복하는 경향을 보였다. 따라서 이들 연구는 중립적 주제로 분류할 수 있다. 이와 같은 연구 결과는 분석의 대상이 되는 중국 학술지 ‘한국어 교수와 연구’에 한정된 것으로 이를 일반화해서 적용할 수 없다는 점을 밝힌다.

결론적으로 본 연구를 통해서 중국 학술지 ‘한국어 교육과 연구’의 학문적 관심이 학문 목적 학습자들을 대상으로 한 교육 관련 연구로 변해 왔다는 것을 확인할 수 있었다. 반대로 문학 관련 연구는 감소하고 있고 동시에 연구자들의 학술 언어 선택이 한국어에서 중국어로 이동하고 있다. 향후 연구에서는 이러한 변화 추세를 조금 더 장기적인 관점에서 분석하기 위해서 자료를 수집하고 이를 분석하여 모니터링을 할 필요가 있다.

현재 토픽 모델링 기법의 발달로 인해서 BERT<sup>12)</sup>와 같은 정교한 분석 도구들이 개발되고 있다. 이러한 도구들은 데이터의 처리 비용이 상대적으로 많이 든다는 문제점이 있으나 향후 개선되어 사용된다면 분석의 정확도를 더 높일 수 있을 것이라 기대된다.

---

12) BERT(Bidirectional Encoder Representations from Transformers)는 구글이 2018년에 발표한 자연어 처리(NLP) 모델이다. BERT는 문장이나 단어의 맥락을 양방향으로 이해하는 것이 특징이다. 이는 전통적인 NLP 모델들이 대개 입력 데이터를 왼쪽에서 오른쪽(또는 그 반대)으로 처리했던 것과 달리, BERT는 양방향(Bidirectional)으로 문맥을 파악한다는 점에서 차별화된다. 따라서 문장의 앞뒤 맥락을 더 잘 반영할 수 있어 자연어 이해 능력이 높다(M. V. Koroteev., 2021). 이를 이용한 한국어 자연어 처리 능력의 개선이 이루어지고 있다(이치훈 외, 2020; 유소엽·정옥란, 2020).

[Abstract]

## Analysis of Research Trends in the Chinese Journal ‘Korean Teaching and Research’

Wang Ruiqi, Kim Yumi

(Seoul School of Integrated Sciences & Technologies)

This study analyzes the changes in language usage and research topics in the articles published in the Chinese journal Korean Teaching and Research from 2016 to 2023. A total of 711 articles were examined, of which 272 written in Chinese were translated by multiple translators and verified for reliability. The articles were collected via the National Knowledge Infrastructure(NKI) website, and the data were analyzed through pre-processing. Keyword frequency analysis, linguistic network analysis, time-series regression analysis, and topic modeling were employed to identify key research topics and their changes over time.

The keyword frequency analysis revealed that translation, education, language, and literature were the core themes. The linguistic network analysis further confirmed the centrality of these themes, showing that they occupied a significant place in the journal and were clustered around key research topics. Topic modeling identified nine major research themes, with education-related topics accounting for the largest proportion. Time-series analysis showed a clear decline in literature-related research over time, while studies on teaching and academic fields showed an increasing trend.

This study highlights the evolving academic focus in Korean language education research in Chinese journals, particularly noting the growing proportion of articles published in Chinese. Future research should monitor these trends and utilize more advanced analytic tools for long-term study.

**Key words** : Korean language education in China, research trend analysis, time-series regression analysis, linguistic network analysis, topic modeling

## [참고문헌]

### □ 기본자료

<https://www.cnki.net> (CNKI 국가지식인프라)

### □ 논문 및 기타 자료

- 김태경·최희련·이홍철, 「토픽 모델링을 이용한 핀테크 기술 동향 분석」, 『한국산학기술학회 논문지』, 한국산학기술학회, 2016, 670-681면.
- 김은실·강승혜, 「한국어 교육 평가 관련 연구 동향 분석 -동시 출현 단어 분석을 중심으로-」, 『한국어 교육』, 국제한국어학회, 2019, 1-20면.
- 김동아, 「한국어 교육에서의 쓰기 피드백 연구 동향 분석 -네트워크 텍스트 분석을 중심으로-」, 『인문사회과학연구』, 부경대학교 인문사회과학연구소, 2022, 85-120면.
- 김민희, 「LDA를 통한 한국어 습득 연구 동향 분석 - 중국어권 학습자 대상 연구를 바탕으로 -」, 『반교어문연구』, 반교어문학회, 2020, 193-213면.
- 박영훈·김창식·송향근, 「토픽 모델링을 활용한 중국인(中國人) 학습자 대상의 한국어(韓國語) 교육(教育) 연구(研究) 동향(動向) 분석(分析)-한국(韓國) 내 학위논문을 중심으로-」, 『어문연구』, 한국어문교육연구회, 2019, 451-474면.
- 박영훈·송향근, 「토픽 모델링을 활용한 한국어 교육의 오류 분석 연구 동향 한국 - 내 학술지 논문과 학위 논문 비교 분석을 중심으로-」, 『Journal of Korean Culture (JKC)』, 한국어문학회학술포럼, 2021, 123-162면.
- 박영훈, 「토픽 모델링을 활용한 한국어 말하기 교육 연구의 동향 분석 -한국 내 연구를 중심으로-」, 『한국어 교육』, 국제한국어교육학회, 2020, 55-82면.
- 박종순·김창식, 「빅데이터 연구 동향 분석: 토픽 모델링을 중심으로」, 『디지털산업정보학회 논문지』, (사) 디지털산업정보학회, 2019, 1-7면.
- 방성원, 「온라인 한국어 교육 관련 연구 동향 분석 - 토픽 모델링과 언어 네트워크 분석을 중심으로 -」, 『한국언어문화학』, 국제한국언어문학회, 2021, 1-30면.
- 신우봉·김일환·김홍규, 「신문 텍스트에서 나타나는 공간명사의 사용 양상과 네트워크 분석」, 『텍스트언어학』, 한국텍스트언어학회, 2010, 221-250면.
- 이수상, 「언어 네트워크 분석 방법을 활용한 학술 논문의 내용분석」, 『정보관리학회지』, 한국정보관리학회, 2014, 49-68면.
- 이수화·김동아, 「네트워크 텍스트 분석을 활용한 대학생의 말하기 및 글쓰기 인식 분석-전남대학교 학생을 대상으로 -」, 『漢城語文學』, 한성어문학회, 2023, 283-319면.
- 이치훈·이연지·이동희, 「사전 학습된 한국어 BERT의 전이 학습을 통한 한국어 기계독해 성능개선에 관한 연구」, 『한국IT서비스학회지』, 한국IT서비스학회, 2020,

83-91면.

- 이정희, 「중국인 학습자 대상 한국어 교육 연구 동향 분석」, 『국어국문학』, 국어국문학회, 2014, 165-197면.
- 유소엽·정옥란, 「BERT 모델과 지식 그래프를 활용한 지능형 챗봇」, 『한국전자거래학회지』, 한국전자거래학회, 2019, 87-98면.
- 원미진, 「중국어권 학습자 대상의 한국어 교육 연구의 동향과 전망」, 『언어사실과 관점』, 연세대학교 언어정보연구원, 2018, 183-204면.
- 왕서기 · 김유미, 「중국에서의 한국어 교육 연구 동향 분석- 학술지 논문(중국조선어문)을 중심으로-」, 『언어사실과 관점』, 연세대학교 언어정보연구원, 2024, 327-359면.
- 정연희, 「토픽 모델링을 활용한 한국어 교육 연구 동향 분석 -한국연구재단의 선정과제를 중심으로-」, 『이중언어학』, 이중언어학회, 2020, 347-367면.
- 조남현·장경완, 「토픽 모델링과 언어 네트워크 분석에 기반한 한국어 발음 연구 동향 분석」, 『학습자중심교과교육연구』, 학습자중심교과교육학회, 2023, 451-468면.
- 조숙균 · 김설연, 「중국어권 학습자를 대상으로 한 한국 국내 한국어교육 연구 동향 분석 -2010년 이후 연구를 중심으로-」, 『반교어문연구』, 반교어문학회, 2022, 265-296면.
- 최일영 · 이용성 · 김재경, 「사회 네트워크 분석에 기반한 도서관 학술DB 이용 패턴 연구: K대학도서관 학술DB 이용 사례」, 『정보관리학회지』, 한국정보관리학회, 2010, 25-40면.
- 최진수 · 정혜원, 「토픽 모델링과 의미연결망 분석을 활용한 영재교육 연구 동향 분석」, 『교육학연구』, 한국교육학회, 2022, 1-28면.
- 허영수, 「동시 출현 단어 분석을 통한 한국어 교육에서의 챗봇 연구 방향 탐색-교육 분야 챗봇 관련 연구 동향 분석을 기반으로-」, 『국제어문』, 국제어문학회, 2023, 641-673면.
- 金银花, 「한국어 문화적 어휘표현에 대한 교육 방법」, 韩国语教学与研究(09).2023, 55-59p.
- 王菲, 「말뭉치를 활용한 정도부사의 사용 양상의 대한 연구」, 韩国语教学与研究(09), 2022, 12-19p.
- 薛娇, 「國外韓國語教學研究進展與評述 ——基於社會網絡分析法和LDA建模」, 韩国语教学与研究(12), 2023, 91-99p.
- 张晶·王平平, 「国内有关韩国语研究动向的分析」, 韩国语教学与研究(01), 2019, 137-144p.
- Blei, D. M., Ng, A. Y., & Jordan, M. I., Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 2003, pp. 993-1022.
- Blei, D. M., Probabilistic Topic Models Surveying a suite of algorithms that

- offer a solution to managing large document archives, *Communications of the ACM*, 55(4), 2012, pp. 77-84.
- Christensen, A. P., & Kenett, Y. N. , Semantic network analysis (SemNA): A tutorial on preprocessing, estimating, and analyzing semantic networks. *Psychological Methods*, 28(4), 2023, pp. 860-879.
- Durbin, J., & Watson, G. S. , Testing for Serial Correlation in Least Squares Regression. I, in *Breakthroughs in Statistics: Methodology and Distribution, Volume 2*, 1992 , New York, NY: Springer New York, pp. 237-259.
- MV Koroteev. , BERT: a review of applications in natural language processing and understanding. eprint arXiv:2103.11943, 2021.
- Mifrah, S., & Benlahmar, E. H. , Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus, *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4), 2020, pp. 5756-5761.
- Griffiths, T. L., & Steyvers, M., Finding scientific topics, *Proceedings of the National academy of Sciences*, 101(suppl\_1), 2004, pp. 5228-5235.
- Röder, M., Both, A., & Hinneburg, A, Exploring the space of topic coherence measures, in *Proceedings of the eighth ACM international conference on Web search and data mining* , February, 2015, pp. 399-408.
- Vayansky, I., & Kumar, S. A. , A review of topic modeling methods. *Information Systems*, 94, 2020.