

COVID-19 팬데믹과 초연결사회의 자화상*

- 대학 내 외국인 유학생 차별을 중심으로 -

강진구**

- I. 머리말
- II. 연구 및 분석 방법
- III. 분석 결과
- IV. 맺음말

국문초록

본 연구는 COVID-19 팬데믹 상황에 놓여 있는 한국 대학생들의 자화상을 살피는 것이다. 이를 위해 필자는 COVID-19 팬데믹으로 인해 트랜스내셔널이 위협받는 상황에서 외국인 유학생에 대한 대학생들의 생각을 텍스트마이닝 기법으로 분석하였다. 주 분석 대상은 ‘에브리타임’에 게시된 외국인 유학생 관련 게시글 중 중앙대와 제주대의 게시글 1,237건이며, 기간은 2019년 12월 1일부터 2021년 8월 31일까지로 한정하였다. 필자는 대학생들의 유학생에 대한 인식을 한국인들의 유학생에 대한 보편적인 인식과 비교하기 위해 네이버뉴스의 유학생 관련 기사의 댓글과 비교 분석도 수행하였다. 수집된 자료는 R프로그램의 “KoNLP”, “WordCloud2”, “ggraph”, “topicmodels” 등의 패키지를 사용해 분석하였으며 그 결과는 다음과 같다.

* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5C2A04082405).

** 제주대학교 탐라문화연구원 학술연구교수

첫째, ‘착짱죽짱’, ‘짱깨’, ‘좃선족’, ‘차이나게이트’, ‘우한폐렴’ 등 한국 사회에 만연해 있는 중국 또는 중국인 유학생에 대한 스테레오타입을 대학생들 역시 반복, 확대, 재생산하고 있다.

둘째, 대학생들은 트랜스내셔널과 초연결성사회라는 세계사적 흐름 속에서 어느새 자신들과 대학이란 공간을 공유하고 있는 유학생의 존재나 변화된 대학 시스템에 대해 동의보다는 반대를, 수용보다는 거부의 태도를 보인다.

셋째, 대학생들은 코로나19 방역조치와 관련한 일부 유학생의 부적절한 행위를 유학생 개인의 문제가 아닌 집단의 속성으로 규정한 후, 그 집단에 대한 극단적인 혐오 감정을 표출하였다.

넷째, 외국인 유학생에 대한 차별행위에는 ‘궤위(闕位: interregnum)의 시대’를 불안과 위협 속에 방치된 채 경쟁으로 내몰린 대학생들의 자기방어의 몸부림이 포함되어 있다는 점에서 특별한 관심을 요한다.

주제어 : 텍스트마이닝, 대학생, 유학생, 에브리타임, 차별, 혐오 표현, 코로나 19, 팬데믹

I. 머리말

오늘날 세계는 ICT의 비약적인 발전과 전세계적인 교통망의 구축, 그리고 자본의 초국적 이동으로 국가 간 경계가 모호한 트랜스내셔널(Transnational)한 상황에 놓여 있다. 게다가 4차 산업혁명의 핵심이라 할 수 있는 이른바 5G(초고속 무선전송 및 접속 기술), 빅데이터, 사물인터넷(IoT), 인공지능(AI) 등 관련 기술의 눈부신 성장은 전 세계 사람들을 네트워크로 거미줄처럼 촘촘하게 연결하는 초연결사회(Hyper-connected Society)를 만들고 있다.

트랜스내셔널과 초연결사회는 21세기를 대표하는 메카트렌드로 등장한 ‘이주와 결합해 전 세계에 많은 변화를 가져오고 있다. 한국 사회 또한 이러한 세계사적 변화의 한 가운데에 있는데, 대학 구성원의 인구

통계학적 변화도 그중 하나이다. 2021년 9월 현재 국내 대학에는 170,380명의 유학생이 재학하고 있다. 국가별로는 62,037명의 베트남이 1위를 차지하고 있고, 한국계 812명을 포함한 중국이 54,041명으로 2위, 우즈베키스탄(9,780), 몽골(8,457), 일본(4,229) 순이다.¹⁾



〈그림 1〉 연도별 유학생수(출처: 한국교육개발원, 2020)

<그림 1>에서 보듯 국내 대학의 유학생 숫자는 코로나19로 인한 팬데믹의 여파로 160,165명(180,131명)²⁾이 재학했던 2019년에 비해 숫자가 줄기는 했지만, 이들은 여전히 대학 교육의 중요 구성원을 이루고 있다.

국내 대학에 재학중인 유학생 수는 경희대(2,939), 성균관대(2,677), 한양대학교(2,429), 고려대학교(2,402), 중앙대학교 서울캠퍼스(2,288) 등 유학생 재학 상위 5개 사립대학은 말할 것도 없거니와 심지어는 전남대(908), 부산대(531), 제주대학교(331) 등 국립대 또한 적지 않은 유학생이 재학하고 있다.³⁾

1) 출입국·외국인정책본부, 『출입국·외국인정책 통계월보 9월호』, 출입국·외국인정책본부, 2021, 31쪽.

2) 한국교육개발원의 통계와 출입국·외국인정책본부의 통계자료 숫자가 다른 것은 유학생을 규정하는 개념상의 차이에서 비롯된다. 다시 말해 개발원은 대학의 학위과정에 재학중인 학생만을 유학생으로 인정하고 있는 반면에 정책본부는 학위과정과 함께 어학연수(한국어 연수)도 유학생에 포함한다.

3) 교육부, 『2020년 국내 고등교육기관 내 외국인 유학생 통계』, 교육부, 2020.

<https://www.moe.go.kr/boardCnts/view.do?boardID=350&lev=0&statusYN=W&s=moe&m=0309&opType=N&boardSeq=82916>(검색일: 2021년 10월 8일)

<그림 1>에서 확인할 수 있듯이 통계를 시작한 2003년 12,314명이었던 국내 재학 유학생 수는 2019년 160,164명으로 약 1,300% 이상 증가하였다. 유학생이 단기간에 급증한 데에는 크게 3가지 요인을 들 수 있다. 하나는 스타디 코리아 프로젝트(Study Korea Project)⁴⁾를 도입해 고등교육의 국제화를 위해 본격적으로 유학생 유치에 나선 정부 정책 때문이다. 두 번째는 학령인구 감소로 인한 대학의 재정난 문제를 해결하고 더 나아가 캠퍼스 구성원의 다양화를 통해 학생들의 국제적 역량 강화를 위해 적극적으로 유학생 유치에 나선 대학들의 노력을 들 수 있다.⁵⁾ 세 번째 요인은 트랜스내셔널과 초연결사회라는 세계사적 흐름에서 형성된 한류의 영향 때문이다. 한류로 대표되는 한국문화에 대한 젊은 외국인들의 관심은 한국을 유학 대상국으로 선택하는 데 중요한 역할을 하였다. 주지하다시피 외국인들의 한국에 대한 인식은 매우 긍정적이다. 문화체육관광부가 발표한 자료에 따르면 한국의 국가이미지는 “한국인이 평가하는 한국에 대한 국가이미지보다 외국인이 평가하는 국가이미지가 높게 나타”⁶⁾날 정도로 좋은 편이다. 다시 말해, 외국인들은 한국과 한국인에 대해 전반적으로 긍정적으로 인식하고 있으며⁷⁾ 특히 한류로 대표되는 한국의 현대문화⁸⁾에 대해서는 직접 경험하기를 원할 정도로 적극적인 관심을 보인다.

4) 스타디 코리아 프로젝트(Study Korea Project)란 고등교육의 국제화를 통한 대학 체제 개선 및 경쟁력 강화를 위해 정부가 마련한 정책이다. 정부는 급격한 국내 입학자원 감소에 따른 대학 미충원 현상을 해소하고, 유학생의 유치를 통한 만성적인 유학수지 적자 개선이라는 목표로 2004년 11월 1단계 스타디 코리아 프로젝트를 가동하였다.

5) 임희진·김제희·정소현·정하영·김현성, 「캠퍼스 국제화에 대한 국내 학생들의 인식 연구」, 『아시아교육연구』 16(1), 서울대학교 교육연구소, 2015, 244쪽.

6) 문화체육관광부 해외문화홍보원, 「2020년 국가이미지 조사 보고서」, 문화체육관광부 해외문화홍보원, 2021, 24쪽.

7) 한국을 포함한 16개국 8,000명을 대상으로 한 조사에서 한국에 대한 긍정은 78.1%로 였으며, 한국인에 대한 호감도는 74.%를 기록하였다(위의 글, 24; 37쪽).

8) 외국인은 한식을 61.5%로 가장 관심이 가는 한국의 현대문화로 꼽았고, “그 다음으로 영화(50.1%) > 현대음악(45.4%) > 드라마(39.6%) > 스포츠(35.2%) > K-뷰티(33.7%)”가 순위를 이뤘다(위의 글, 21쪽).

유학생 증가로 인해 대학가의 모습은 이른바 ‘캠퍼스의 국제화’라는 새로운 풍경을 형성하고 있다. 캠퍼스 곳곳에는 외국인 유학생을 위한 각종 멘토성 프로그램에 대한 안내 현수막이 게시되어 있고, 학교의 공식 홈페이지나 대학생들이 주로 활용하는 교내 온라인 커뮤니티 등에도 유사한 글들이 게시되고 있다. 이같은 풍경은 교육부와 대학이 처음 유학생을 유치하면서 기대했던 고등교육의 국제화를 통한 국가 브랜드 가치 제고와 대학의 재정난 해소, 그리고 학생들로 하여금 새로운 세계와 문화, 언어를 습득하게 해 대학 교육을 더욱 풍요롭게 하겠다는 목표가 외형적으로 성공한 것처럼 보이게까지 하였다.

하지만, 이른바 ‘캠퍼스 국제화’ 정책은 “캠퍼스 국제화를 통해 주어지는 긍정적인 부분보다 국제화의 이상과 현장에서의 괴리, 불편, 또는 상대적 박탈감과 같은 부정적인 측면이 더 크게⁹⁾ 부각되고 있다. 유학생에 관한 선행연구¹⁰⁾에 따르면 국내에 재학중인 외국인 유학생들은 초국적 이동성을 경험한 유학생 일반이 경험하는 수준을 훨씬 상회하는 어려움에 직면해 있다. ‘유학생 차별’이 그것이다. 다시 말해, 국내 대학에 재학중인 유학생들은 유학생 일반이 경험하는 어려움에다가 한국인 대학생들에 의해 반복적이고 지속적으로 행해진 직·간접적인 인종차별 내지 ‘유학생 차별’이라는 이중의 고통을 겪고 있는 것이다.

유학생들은 익숙한 환경을 떠나 타국에서 생활하기 때문에 “언어문제, 경제문제, 사회적응문제, 향수병, 개인적 역할갈등”¹¹⁾ 등에서 개인간의 차이가 있기는 하지만 어려움을 겪는다. 국내 유학생들은 여기에 더해 대학 생활 전반에서 보이지 않게 작동하는 미세 차별은 물론이고 신인종주의에 근거한 무차별적 집단 차별 등으로 극심한 스트레스에

9) 임희진 외, 앞의 글, 262쪽.

10) 대표적인 연구로는 김정숙(2021), 진사선(2019), Desvinta Ayu Iriani, Ito Hiroko (2018), 임운서 외(2016), 임희진 외(2015), 강운선(2016), 김종태 외(2013), 김해연 외(2019), 황동진 외(2021), 김세경(2017), 김정희 외(2018), 백근영(2020), 민예슬 외(2018) 등이 있다.

11) Desvinta Ayu Iriani, Ito Hiroko, 「대구경북지역 거주 아시아권 출신 유학생의 학교 내 차별경험 출신국가지역에 따른 위계적 차별」, 『한국사회학회 사회학대회논문집』, 한국사회학회, 2018, 633쪽.

노출되고 있다. 중앙대 학생생활상담연구소에 따르면 중앙대에 재학중인 외국인 유학생들은 유학생회에 51.0%가 만족¹²⁾하는 편이지만, 약 31.4%의 유학생은 ‘교수와 학생들의 차별’ 때문에 학교생활이 만족스럽지 못하며, 특히 중국 유학생의 경우 학교생활을 하면서 약 26.4%가 직접적으로 차별을 받았다고 밝히고 있다.¹³⁾

그런데 유학생과 관련한 불만족은 인종차별과 미세차별에 노출된 유학생뿐만 아니라, 그들과 함께 생활하는 한국인 학생들 사이에서도 표출된다. 한국인 학생들은 “학교 수업이나 조모임, 학과행사 등 다양한 장면”¹⁴⁾에서 차별을 당하고 있다는 유학생들의 주장에 대해 피해자는 오히려 자신이라며 강력하게 반발한다. 한국인 학생들은 유학생으로 인해 강의의 질이 떨어지고 있으며, 조별 과제 등에서는 자신들이 손해를 보고 있다고 생각하기까지 한다. 심지어는 역차별을 당하고 있다는 주장과 함께 유학생에 대한 노골적인 인종차별적 행위를 자행하기까지 한다.

유학생에 대한 인종차별로 학내 문제가 된 대표적인 사례는 중앙대학교에서 발견된 ‘유학생 혐오 낙서 사건’(2017.10.26.)을 들 수 있다. 문제가 된 낙서에는 중국 국가주석과 중국 유학생을 비하하는 발언이 적혀 있었다. 해당 사건을 접수한 중앙대 측은 교내 인권센터의 활동을 강화하는 한편, 이례적이라 할 만큼 신속하게 ‘차별행위에 관용 없다’라는 총장의 특별 성명까지 발표하며 유학생에 대한 차별 문제를 차단하려 했다. 하지만 이 같은 학교측의 노력에도 불구하고 학교의 공식적인 온라인 커뮤니티인 ‘중앙인’에 유학생을 차별하는 내용의 글이 버젓이 게시되는 사건(2018.9.18.)이 발생하는 등 유학생에 대한 차별과 이로 인한 학내 갈등 문제는 아직도 계속되고 있다.¹⁵⁾

12) 중앙대학교 학생생활상담연구소, 「2017 유학생 의식조사 및 정신건강실태 보고」, 중앙대학교 학생생활상담연구소, 2017, 13쪽.

13) 위의 글, 34쪽.

14) 민예슬·김창대, 「외국인 유학생의 일상적 차별경험이 매일의 우울에 미치는 영향: 내부귀인의 조절효과」, 『상담학연구』 19(6), 한국상담학회, 2018, 114쪽.

15) 자세한 내용은 「혐오 글에 유학생 빨났다」, 《중대신문》(2018.10.8.)을 참조.

문제의 요체는 유학생에 대한 차별과 이로 인한 갈등이 비단 중앙대학교만의 문제가 아니라는 데 있다. 유학생에 대한 비하와 차별은 정도의 차이는 있지만 대학 사회 전체에 만연되어 있다고 봐야 한다. 김정숙에 따르면 대학 관련 정보 공유를 목적으로 제작된 대학교 커뮤니티인 ‘에브리타임’에는 각종 혐오 발언들이 범람하고 있는데, 그중 하나가 “외국인 유학생의 언어와 수학능력에 대한 고정관념과 차별 발언”¹⁶⁾이다. 다시 말해, ‘에브리타임’에는 익명의 막말과 비방, 소수자 혐오, 그리고 음란 표현과 정치적으로 편향된 표현 등이 적지 않게 게시¹⁷⁾되고 있으며, 특히 코로나19로 인한 팬데믹 상황에서는 중국인에 대한 원색적인 비난이 표출되기도 하였다.

이 논문의 목적은 코로나19로 인한 팬데믹 상황에서 외국인 유학생에 대한 한국사회의 일반적인 인식을 살펴보고, 이러한 인식이 실제 캠퍼스에서 자신들과 함께 생활하는 유학생들을 바라보는 학생들의 인식과 어떠한 차이점을 갖는지를 파악하는 것이다. 이를 위해 필자는 2019년 12월 1일부터 2021년 8월 31일까지 유학생 관련 ‘네이버 뉴스’ 기사와 댓글을 분석하였다. 아울러 대학생의 인식이 날것 그대로 표출되고 있는 ‘에브리타임’에 게시된 유학생 관련 글과 댓글 또한 텍스트마이닝(Text Mining) 기법을 활용해 분석하였다. 필자가 유학생과 관련한 뉴스 기사와 댓글을 분석 대상으로 삼은 이유는 이들 텍스트가 “다른 사람들의 인식 정도와 사회적인 여론의 온도를 감지하고 세상을 이해하는 척도”¹⁸⁾로 작용할 수 있다는 판단 때문이었다. 이 분석을 통해 유학생에 대한 한국인들의 인식과 감정뿐만 아니라, 그러한 인식이 어떠한 연관속에서 형성되는지를 밝힐 수 있을 것이다. 또한 코로나19로 인한 팬데

16) 김정숙, 「초국적 주체와 COVID-19: 유학생의 인종언어 이데올로기 차별경험, 정체성, 그리고 행위주체성」, 『현대사회와다문화』 11(2), 대구대학교 다문화사회정책연구소, 2021, 82쪽.

17) 이지, 「대학생의 필수 앱 ‘에브리타임’의 명과 암」, 《소비라이프》, 2021.

<http://www.sobilife.com/news/articleView.html?idxno=30076>

18) 권상희·김익현, 「온라인 댓글 인식과 댓글 활동의 관계에 관한 연구: 댓글의 신뢰도와 인터넷뉴스 수용자의 수용경향을 중심으로」, 『한국언론정보학보』 42(2), 한국언론정보학회, 2008, 47쪽.

믹 상황에서 유학생에 대한 차별이 존재한다면, 차별의 실체 또한 시각적인 표현을 통해 명확하게 드러낼 수 있을 것이다.

Ⅱ. 연구 및 분석 방법

1. 분석 대상 자료 수집

유학은 본질적으로 지리적 이동을 전제로 한다. 따라서 유학생은 필연적으로 트랜스내설한 문화체험을 경험하게 된다. 이러한 유학 또는 유학생이 지닌 독특한 특징 때문에 유학생들은 초국적 주체로서 “근대적 국가와 민족에 결박되고 정체”된 존재이기보다는 “유동적이고 역동적인” 존재로 표상되었다.¹⁹⁾

하지만, 코로나19로 인한 팬데믹으로 국가 간의 연결성은 어느덧 ‘봉쇄’와 ‘입국금지’, 그리고 ‘격리조치’ 등으로 대체되고 있으며 트랜스내설한 문화체험의 주체로서의 유학생의 위치 또한 흔들리고 있다. 게다가 외국인 유학생을 가시적인 보균자로 보는 ‘우한폐렴’ 등과 같은 ‘질병의 타자화’²⁰⁾가 진행되고 있는 상황에서 한국인들과 대학생들은 외국인 유학생을 어떻게 인식하고 있을까? 필자는 코로나19가 발생한 이후 외국인 유학생과 관련한 ‘네이버 뉴스 기사와 댓글을 수집하였다. 분석 대상 기사 범위는 외국인 유학생과 관련된 주요 뉴스 기사로 한정했으며, 검색 기간은 2019년 12월 1일부터 2021년 8월 31일까지로 한정하였다. 댓글 수집을 위한 프로그램은 파이썬 언어를 사용해 직접 개발하였다. 1차로 수집된 뉴스 기사는 모두 642건이었고, 거기에 달린 댓글 수는 모두 18,075건이었다. 한편 ‘에브리타임’의 경우 중앙대학교와 제주대학교 2곳을 선정하여 게시글과 그에 따른 댓글을 모두 수작업으로 수집

19) 김정숙, 앞의 글, 94쪽.

20) ZHOU NAN, 「중국과 관련된 코로나-19 뉴스이용이 중국인에 대한 제노포비아 정서에 미치는 영향」, 서강대 석사학위, 2020, 17쪽.

하였다. 1차로 수집된 자료는 중앙대가 110건의 게시글과 845건의 댓글이었고, 제주대는 게시글 64건과 댓글 218건 이었다.

필자는 먼저 수집된 유학생 관련 뉴스를 댓글이 많이 달린 수를 기준으로 분류하여, 당시 독자들이 어떤 기사에 관심을 갖고 있었는지를 파악하고자 했다. 642건의 유학생 관련 뉴스는 댓글이 많은 순서를 기준으로 내용상으로 분류하면 대략 5가지로 분류할 수 있다. ①코로나19 확진을 받고 무료로 치료받는 중국인 확진자와 관련해 외국인 의료보험 문제를 다룬 기사, ②격리지를 무단 이탈한 베트남 유학생을 비롯한 격리 이탈자에 대한 기사, ③베트남 유학생에 대한 성범죄를 저지른 경찰관에 대한 기사, ④코로나19와 팬데믹 상황에서 유학생 대거 입국으로 인한 감염 공포를 다룬 기사, ⑤유학생에 대한 차별과 역차별을 주장하는 한국인 학생들과의 갈등을 다룬 기사가 그것이다. 댓글이 가장 많이 달린 기사는 [월 7만원 내고 4억7500만원 치료받은 중국인, 건보 급여 어찌나(머니투데이)로 유효 댓글만 4,597건이 달렸다. [베트남 유학생들 격리지 이탈...강제 추방 위기]라는 기사가 1,172건의 유효 댓글이 달려 2위에 위치해 있고, 3위는 유효댓글 1,000개를 기록한 [19살 베트남 유학생 알바 첫 날, 그녀를 덮친 건 전직 경찰이었다]라는 기사가 차지했다.

필자는 이들 기사를 포함해 모두 642건의 유학생 관련 기사에 달린 18,075건의 댓글을 분석하였다.

2. 분석 방법

텍스트마이닝 기법으로 텍스트를 분석하기 위해서는 먼저 수집된 데이터를 정제해야 한다. 필자는 한글 자연어 처리 패키지인 KoNLP패키지를 통해 전처리를 하였는데, 이 과정에서 분석에 필요한 필자만의 사용자 정의 사전과 불용어 사전을 추가하여 분석하였다.

구체적인 분석 내용으로는 댓글 전체를 대상으로 한 빈도 분석(워드 클라우드)과 토픽 분석, 감정 분석, 의미망 분석 등을 실시하였다.

Ⅲ. 분석 결과

1. 빈도분석

코로나19 팬데믹 상황에서 유학생에 대한 한국인들의 인식은 관련 뉴스에 달린 댓글에 대한 빈도를 분석해 보면 쉽게 유추할 수 있다.

〈표 1〉 ‘유학생’ 관련 뉴스 댓글 상위 50개 단어

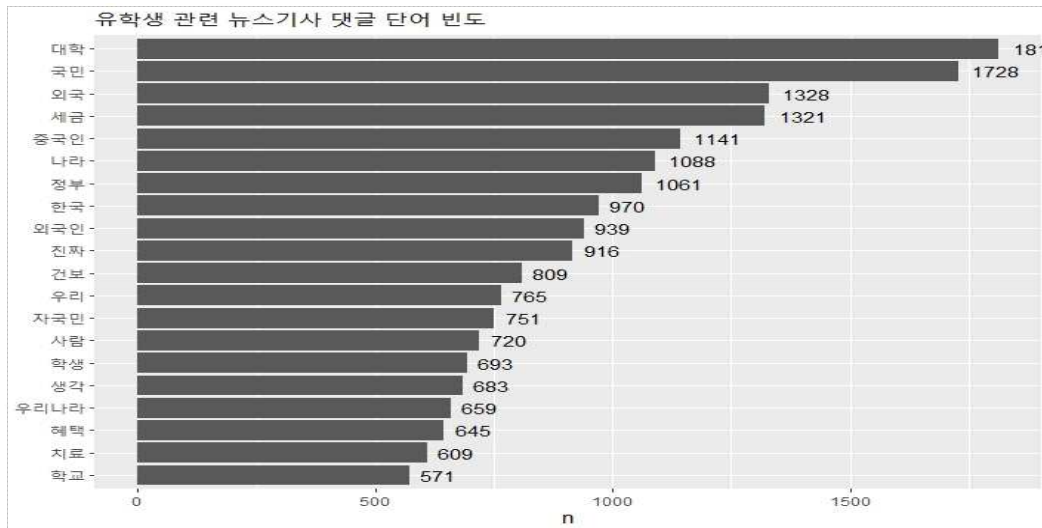
순위	단어	빈도수	순위	단어	빈도수
1	대학	1813	26	미국	468
2	국민	1728	27	입국	468
3	외국	1328	28	유학생	464
4	세금	1321	29	건강	432
5	중국인	1141	30	짱깨	418
6	나라	1088	31	경찰	411
7	정부	1061	32	기사	402
8	한국	970	33	문재양	378
9	외국인	939	34	마스크	364
10	진짜	916	35	호구	356
11	건보	809	36	국가	345
12	우리	765	37	지원	341
13	자국민	751	38	한국인	336
14	사람	720	39	대통령	332
15	학생	693	40	정권	306
16	생각	683	41	누구	298
17	우리나라	659	42	병원	295
18	혜택	645	43	교수	282
19	치료	609	44	교육	268
20	학교	571	45	국적	266
21	문재인	566	46	정치	263
22	대한민국	559	47	지방	261

23	문제	549	48	격리	250
24	보험	528	49	만원	249
25	등록금	509	50	서울	249

위의 <표 1>은 빈도 분석을 통해 찾아낸 네이버 뉴스의 ‘유학생’ 관련 댓글의 주요 키워드 상위 50개이다. 검색 키워드인 ‘외국인 유학생’을 제외하고 주요 출현 단어를 통해 댓글의 내용을 살펴보도록 하자. 최상위 출현빈도는 모두 1,000회 이상 빈도수를 기록한 어휘들로 ‘대학’, ‘국민’, ‘세금’, ‘중국인’, ‘나라’, ‘정부’ 등이 차지하고 있다. 그 다음으로는 ‘한국’, ‘진짜’, ‘건보’, ‘자국민’, ‘혜택’, ‘치료’, ‘문재인’, ‘보험’, ‘등록금’ 등의 단어가 500회 이상 높은 출현 빈도를 보이고 있다. 그리고 이들 단어보다 출현 빈도는 적지만 ‘짱깨’, ‘문재양’, ‘호구’, ‘격리’ 같은 어휘들이 상위 50위에 출현하고 있음을 볼 수 있다.

빈도수 분석을 통해 유학생들에 대해 한국인들의 인식을 유추할 수 있는데, 최상위 빈도수를 차지한 단어들에서 한국인들은 유학생을 트랜스내셔널과 초연결사회라는 세계사적 흐름이 산출한 새로운 문화적 다양성의 실질적 존재나 또는 한국사회(대학)의 국제화, 개방화 등을 파악하는 기준으로 삼기보다는 국민의 세금만을 축내는 부정적 존재로 인식하고 있음을 살필 수 있다. 또한, ‘한국’, ‘진짜’, ‘건보’, ‘자국민’, ‘혜택’, ‘치료’, ‘문재인’, ‘보험’, ‘등록금’ 등의 어휘를 통해서 코로나19 팬데믹 상황임에도 불구하고 외국인 유학생에 대해 입국금지 등의 조치를 취하지 않고, 게다가 코로나19에 확진된 외국인마저 치료해주는 문재인 정부의 코로나19 대응 정책에 대한 비판의식 또한 어렵지 않게 찾을 수 있다.

‘짱깨’, ‘문재양’, ‘호구’, ‘격리’ 등의 어휘는 코로나19로 인한 팬데믹이라는 비상 상황에서 유학생 관련 기사를 접한 한국인들의 감정이 유학생은 물론이고 문재인 정부에 대해서도 매우 부정적임을 엿보게 한다.



〈그림 2〉 유학생 관련 뉴스 기사 댓글 단어 빈도 Top20

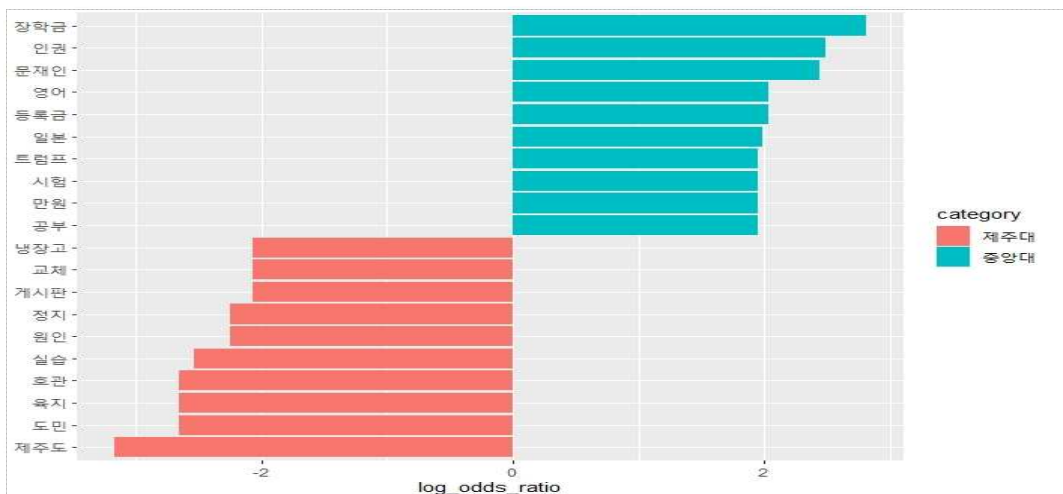
한편, 필자는 ‘에브리타임’의 게시글도 빈도분석을 실시하였는데, ‘생각’, ‘유학생’, ‘진짜’, ‘사람’, ‘중국인’ 등의 어휘가 최상위 빈도를 보이고 그 다음이 ‘학교’, ‘문제’, ‘학생’, ‘우리나라’, ‘친구’, ‘장학금’, ‘조선족’ 등의 어휘가 출현하고 있다. 이러한 단어들은 ‘에타’의 게시글과 댓글들이 코로나19 팬데믹이라는 비상 상황임에도 불구하고 여전히 자신들의 편의만을 좇아 생각 없이 행동하는 중국인 유학생에 대한 비판을 중심에 놓고, 그 주위로 중국인 유학생을 비롯한 유학생 문제를 적절히 관리하지 못한 대학 당국과 외국인 유학생에 대해 입국금지 조치를 취하지 않은 문재인 정부에 대한 비판이 배치되어 있음을 보여준다.

또한 ‘진짜’, ‘조선족’, ‘짱깨’, ‘인권’ 등의 어휘 추출을 통해 코로나19 팬데믹 상황에서 ‘에타’에 글을 작성하는 대학생들의 감정 역시 앞서 살펴본 한국인들의 인식처럼 매우 부정적이고 분노에 차 있음을 엿보게 한다. 이상의 빈도수 분석은 시각화를 통해 더욱 직관적으로 확인할 수 있다.



<그림 3> '에브리타임'의 유학생 관련 게시글과 댓글을 기반으로 한 워드클라우드

<그림 3>은 '에타'의 게시글과 댓글의 빈도 분석 결과를 워드클라우드 방식을 통해 시각화 한 것이다. 필자는 '에브리타임'의 유학생 관련 게시글과 댓글을 R 프로그램의 'worldcloud2' 라이브러리로 상위 출현 단어 100개를 워드클라우드로 구성하였다. 워드클라우드에서는 빈도수가 높은 단어가 굵고 큰 글씨로 나타나는데, 이를 통해 대학생들의 관심사를 시각적으로 확인할 수 있다.



<그림 4> '에브리타임'의 주요 단어(중앙대 VS 제주대)

<그림 4>는 ‘에타’ 게시글과 댓글을 빈도 분석해 추출된 단어들을 중앙대와 제주대로 구분한 후 ‘로그 오즈비’²¹⁾ 분석을 통해 상대적으로 중요한 단어를 10개씩을 뽑아 그래프로 나타낸 것이다. 중앙대에 달린 글에서는 ‘장학금’, ‘인권’, ‘문재인’, ‘영어’, ‘등록금’ 등의 어휘가 제주대의 글보다 많이 사용되고 있다. 제주대의 글에는 ‘제주도’, ‘도민’, ‘육지(생)’, ‘호관(6호관-기숙사 명칭)’ 등의 어휘가 상대적으로 중요하게 등장한다. 이것은 중앙대에 달린 댓글들이 제주도에 달린 댓글보다 유학생 일반에 대한 문제의식이 강하며, 제주대의 경우 상대적으로 지역적 특성을 반영한 글들이 많이 게시되고 있음을 보여준다.

2. 토픽 분석

코로나19로 인한 팬데믹 상황에서 유학생 관련 뉴스에 달린 댓글들의 내용은 무엇일까?

- ①문재인이 중국인 특혜로 중국공산당에 보은한다고 나는 본다.(4,396)
- ②아니 장난해?? 당연히 대학에서 뺏어내야지 무슨 세금이 넘치고 넘치냐?? 어느 사고 방식이야(4,221)
- ③3년정도 거주했다고 외국인에게 투표권을 주는 나라...국민은 부동산 폭등에 내집마련의 꿈이 사라지고 있는데 문재인 정부의 규제 밖인 중국인들은 한국의 부동산을 쓸어담고 있으니...(3,462)
- ④얼굴공개바람(3,436)
- ⑤완전 어이없음 중국유학생 받아서 보호하고 우리나라 학생 내쫓고?

21) 오즈비(odds ratio)는 어떤 사건이 A 조건에서 발생할 확률이 B 조건에서 발생할 확률에 비해 얼마나 더 큰지를 나타내는 값이다. 이 분석을 통해 중앙대의 게시글과 댓글에는 많이 사용되었지만, 제주도의 게시글과 댓글에는 적게 사용된 단어-상대적으로 중요한 단어-추출을 통해 이 두 집단의 댓글의 차이를 보다 선명하게 파악할 수 있다(김영우, 『Do it! 쉽게 배우는 R 텍스트 마이닝』, 이지스퍼블리싱, 2021, 77쪽).

무슨 이런 개엇같은 상황이 중국 유학생들을 못들어오게 하는게 맞지!! 진짜 열통 터져서 화병나겠네(2,943)

인용문은 유학생 관련 뉴스 기사 댓글 중에서 독자들의 공감을 가장 많이 받은 댓글 상위 5개이다. ①의 댓글은 [‘無자격’ 비자 중국인, 아파트 3채 사서 월세 160만원 챙겼다]라는 기사에 달린 981개의 댓글 중 하나이고, ②의 댓글은 [대학등록금 반환에 국민세금 들어가야 하나]라는 기사에 대한 680개의 반응 중 하나이다. ③의 댓글은 ①의 기사에 달린 댓글 981개 중 두 번째로 추천을 많이 받은 댓글이고, ④는 [19살 베트남 유학생 알바 첫 날, 그녀를 덮친 건 전직 경찰이었다]라는 기사에 달린 1,000개 중 하나이며, 마지막으로 ⑤의 댓글은 [“100만원이 넘는 기숙사비 차액을 어디서 구하나요”...한국 학생들 ‘울상’]의 기사에 달린 477개의 댓글 중 하나이다.

댓글에는 외국인 유학생에 대해서는 말할 것도 없거니와 유학생 유치 정책을 실시하고 있는 대학과 정부에 대한 한국인들의 솔직한 생각이 드러나 있다. 먼저 동일한 기사에 달린 댓글인 ①과 ③은 국민들은 문재인 정부의 잘못된 부동산 정책으로 인한 부동산 가격 폭등 때문에 내집 마련의 꿈을 상실한 채 고통받고 있는데, 정부는 국민을 삶에 비웃기라도 하듯이 중국인에게는 각종 혜택을 주고 있다고 비판한다. 더불어 이러한 ‘중국인 우대 한국인 차별’ 정책은 모두 중국공산당에 보은하려는 문재인 대통령 때문이라는 인식이다. 그런데 이 댓글에는 ‘중국공산당’이란 부정적인 어휘가 ‘보은’이란 긍정적인 어휘와 절합관계를 형성함으로써 문재인 정부의 부동산 정책에 대한 강력한 비판의식을 표출하고 있다.

댓글 ②와 ⑤는 유학생 관련 정부의 정책에 대한 비판보다는 유학생을 유치한 대학에 대한 반감을 드러내고 있다. ②의 경우는 코로나19와 팬데믹 상황으로 인해 유학생이 감소하여 대학의 재정난이 가속화된 상황에서 온라인 강의로 인한 학생들의 등록금 인하 요구와 관련한 대학들의 입장을 제시한 기사의 댓글이다. ⑤는 각 대학들이 입국 유학생

들의 격리 공간으로 대학 기숙사를 사용한 것에 대한 불만과 기숙사 배정 등에서 상대적으로 유학생에 비해 역차별을 당하고 있다는 분노를 표출하고 있다.

이처럼 댓글을 꼼꼼하게 읽으면 댓글의 내용을 온전히 파악할 수 있다. 하지만 인용문은 분석대상 댓글 18,075건 중 5개에 불과하다. 그러므로 18,075건의 댓글을 이처럼 읽으면서 분석하는 것은 어렵다.

토픽 분석은 대규모의 문서가 어떠한 주제를 구성하고 있는지를 확률/통계적인 접근법으로 분석하는 기법이다. 일반적으로 같은 주제를 이루는 단어들은 동일한 문장 혹은 문서 내에 함께 등장하기 마련인데, 이를 기계적으로 분석하면 자주 쓰이는 단어들의 집합을 얻을 수 있고 이를 통해 문서의 토픽을 도출할 수 있다.²²⁾

이번 연구에서는 “topicmodels” 패키지가 제공하는 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 기법을 사용하였다. 분석 대상은 앞서 뉴스 검색으로 수집한 전체 댓글 18,075건이다.

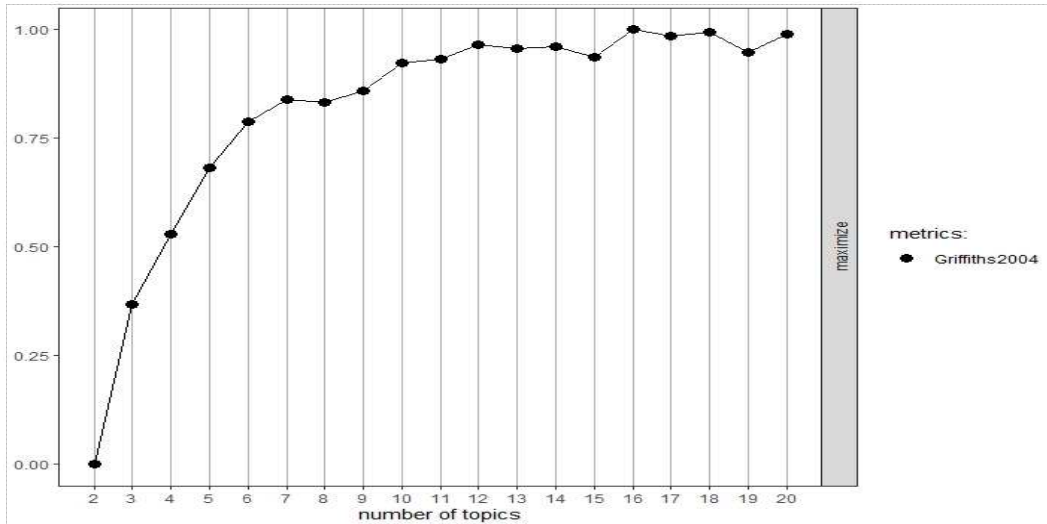
토픽모델링 분석은 빈도수 분석에서는 찾아내기가 쉽지 않은 텍스트의 의미 군집을 탐색한다는 점에서 텍스트마이닝 분석에서 자주 사용하는 기법이다. 하지만 이 분석은 최적의 결과를 도출하기 위해서는 연구자가 토픽 개수를 정해줘야만 한다.²³⁾ 필자는 토픽 분석에 작동하고 있는 자의성을 줄이기 위해 “ldatunning”패키지의 FindTopicsNumber함수를 이용해 적정 K 값을 산출하려 했다²⁴⁾.

필자는 FindTopicsNumber함수를 사용해 분석 데이터의 K값을 산출했는데, 그 결과는 <그림 5>와 같다.

22) 강진구, 「팬데믹과 혐오-베트남 관련 기사의 댓글을 중심으로」, 『다문화콘텐츠연구』 37, 중앙대학교 문화콘텐츠기술연구원, 2021, 201쪽.

23) 강진구, 「빅데이터 분석을 활용한 다문화 연구 동향 분석」, 『다문화콘텐츠연구』 35, 중앙대학교 문화콘텐츠기술연구원, 2020, 65쪽.

24) 백영민, 『R를 이용한 텍스트 마이닝』 개정판, 한울, 2020, 286쪽.



〈그림 5〉 잠재토픽 개수 산출 그래프

〈그림 5〉는 FindTopicsNumber 함수가 찾아낸 적정 K 값을 그래프로 표현한 것이다. 그래프에서 볼 수 있듯이 ‘Griffiths2004’ 선이 토픽 수가 7이 될 때까지 X축인 성능 지표가 상승하다가 이후 완만한 모습을 보인다. 성능이 비슷한 경우 단순한 모델을 사용하는 것이 해석 가능성을 높이기 때문에 K 값으로 7이 적당하다.²⁵⁾ 필자는 R프로그램의 “topicmodels” 패키지를 활용해 7개의 중심 토픽을 <표 2>와 같이 산출하였다.

〈표 2〉 ‘유학생’ 관련 기사의 댓글 토픽분석 결과

중요도(순위)	토픽 1	토픽 2	토픽 3	토픽 4	토픽 5	토픽 6	토픽 7
1	입국	중국인	나라	외국인	경찰	학생	한국
2	유학생	문재인	진짜	건보	인간	학교	문제
3	정부	정부	우리	혜택	자기	등록금	미국
4	중국인	대한민국	사람	치료	공무원	생각	생각
5	마스크	짱깨	우리나라	보험	사회	교수	국가
6	입국금지	문재양	자국민	건강	오지	교육	지원
7	금지	정권	기사	호구	쓰레기	대학교	한국인

25) 강진구(2021), 앞의 논문, 202쪽.

8	코로나	대통령	기자	자국민	세상	수업	필요
9	개학	민주	사람들	병원	어디	지방	국적
10	격리	조선족	누구	의료보험	강제	서울	세계

한편, 토픽모델링을 활용한 토픽 분석 시에는 빈도가 너무 높은 단어들은 모든 토픽에 등장해 개별 토픽의 특징 파악을 어렵게 한다. 필자는 빈도수가 1,000회 이상인 어휘를 제외하고 분석하였다.

<표 2>는 7개의 토픽으로 분류하여 각각의 토픽에 등장하는 상위 단어 10개를 표로 나타낸 것이다. 주요 토픽의 내용은 다음과 같다.

먼저, ‘입국’, ‘유학생’, ‘정부’ 등의 어휘가 상위 단어로 출현한 <토픽 1>은 ‘중국인’, ‘마스크’, ‘입국금지’, ‘금지’, ‘코로나’ 등의 단어와 연관을 맺음으로써 코로나19 확산에 따른 중국인 유학생의 입국금지를 둘러싼 일련의 논의들이 중심 주제를 형성하고 있음을 보여준다.

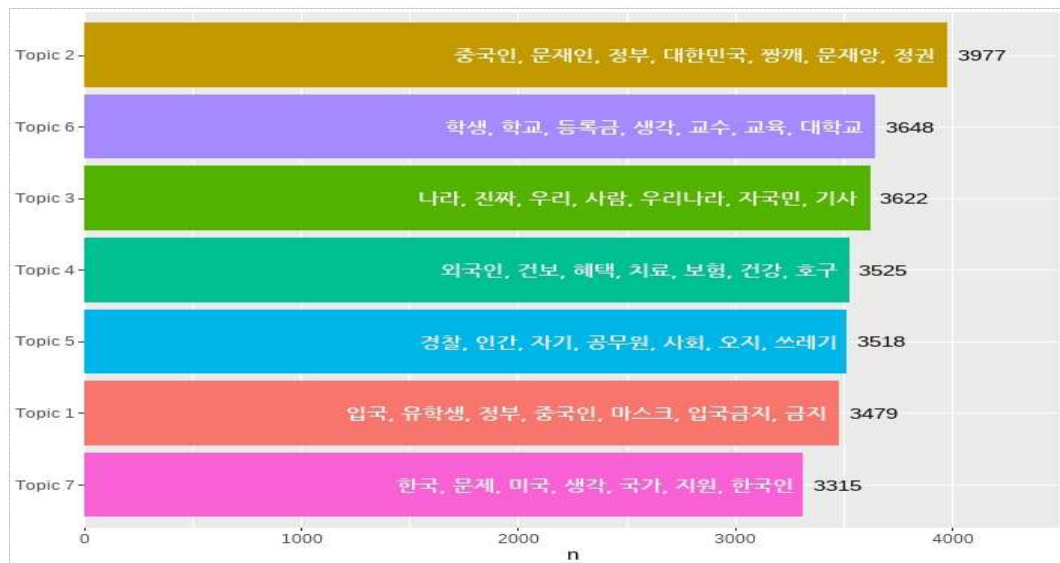
‘중국인’, ‘문재인’, ‘정부’, ‘대한민국’ 등이 앞에 출현한 <토픽 2>는 ‘짱깨’, ‘문재양’, ‘민주당’, ‘조선족’ 등의 어휘와 연관을 맺음으로써 코로나19 팬데믹 상황에서 중국인 유학생의 입국을 허용한 문재인 정부와 중국인에 대한 혐오가 하나의 주제를 이루고 있다. <토픽 3>은 상위에 위치한 ‘나라’, ‘진짜’, ‘우리’ 등의 단어가 ‘사람’, ‘우리나라’, ‘자국민’, ‘기사’, ‘기자’ 등과 연관을 맺음으로써 코로나19 팬데믹 상황과 관련한 언론 관련 이슈가 주요한 담론을 형성하고 있음을 보여준다. ‘외국인’, ‘건보’, ‘혜택’ 등이 등장한 <토픽 4>는 ‘치료’, ‘보험’, ‘건강’, ‘호구’, ‘병원’, ‘의료보험’ 등의 단어와 연관을 맺고 있어 코로나19 확진 외국인에 대한 의료보험 적용 문제가 명시적으로 논의되고 있다.

<토픽 5>는 ‘경찰’, ‘인간’, ‘자기’, ‘공무원’ 등의 어휘가 ‘쓰레기’, ‘세상’, ‘강제’ 등의 단어와 관계를 맺음으로써 경찰관에 의한 베트남 유학생에 대한 강간을 둘러싼 문제가 중심을 이루며, <토픽 6>은 ‘학생’, ‘학교’, ‘등록금’ 등 상위에 위치한 단어들이 ‘교수’, ‘교육’, ‘대학교’, ‘수업’, ‘지방’, ‘서울’ 등의 어휘와 연관을 맺음으로써 학령인구 감소에 대한 대학 위기와 그 극복 방안으로서의 유학생을 유치한 대학사회의 현

실 등이 주요 토픽을 형성하고 있다.

마지막으로 ‘한국’, ‘문제’, ‘미국’, ‘생각’이 중요 어휘로 등장하는 <토픽 7>은 ‘국가’, ‘지원’, ‘필요’, ‘국적’, ‘세계’ 등의 어휘와 결합하여 코로나19 팬데믹 상황에서 유학생 문제를 바라보는 한국과 미국의 차이 등이 논의의 중심을 이루고 있다.

한편, 토픽 분석은 개별 댓글들을 토픽별로 분류하는 작업을 수행하기도 한다. 즉 R 프로그램의 tidytext 패키지의 tidy함수는 이른바 ‘문서별 토픽 확률’인 감마(gamma)값을 추출하게 되는데, 이를 이용하면 문서를 확률이 가장 높은 토픽으로 분류할 수 있다.²⁶⁾



<그림 6> 문서별 토픽 수와 주요 단어

<그림 6>은 댓글들이 각각의 토픽으로 분류된 수와 각 토픽의 주요 단어를 그래프로 나타낸 것이다. 댓글들은 ‘문서별 토픽 확률’에 의해 최소 3,315건에서부터 최대 3,997건씩 7개의 토픽으로 분류되고 있다. 문서별 토픽 수 분석은 외국인 유학생 관련하여 ‘네이버 뉴스’ 댓글의 전체적인 흐름을 파악하게 한다. <그림 6>을 통해 우리는 외국인 유학생 관련 댓글들이 코로나19 팬데믹 상황에서 한국과 미국의 유학

26) 김영우, 앞의 책, 210-214쪽.

생 정책의 차이 등에 대한 이슈나 중국인 유학생 입국금지 등의 문제보다 <토픽 2>의 중국인 유학생에 대한 직접적인 혐오 표현과 <토픽 6>의 유학생 유입 정책으로 변화된 대학사회에 대해 상대적으로 더 많은 관심을 갖고 있음을 확인할 수 있다.

3. 감정 분석

코로나19 팬데믹 상황에 직면한 대학생들은 자신들과 함께 생활하는 외국인 유학생에 대해 어떤 감정을 갖고 있었을까? 감정 분석은 댓글 작성자가 어떤 감정으로 글을 작성했는지를 파악하게 해주며, 특정한 이슈에 대해 어떠한 감정을 갖고 있는지도 파악할 수 있게 한다.²⁷⁾ 여기서는 ‘에타’에 게시된 글들이 어떤 생각을 담고 있는지를 분석하기 위해 수집된 글 전체를 대상으로 감정 분석을 시도하였다. 일반적으로 댓글에는 거친 언어와 정제되지 않은 욕설, 그리고 인신공격과 혐오 표현들을 담고 있고, 긍정적인 주장보다는 부정적인 견해가 훨씬 강력하게 표출되어 있다.²⁸⁾

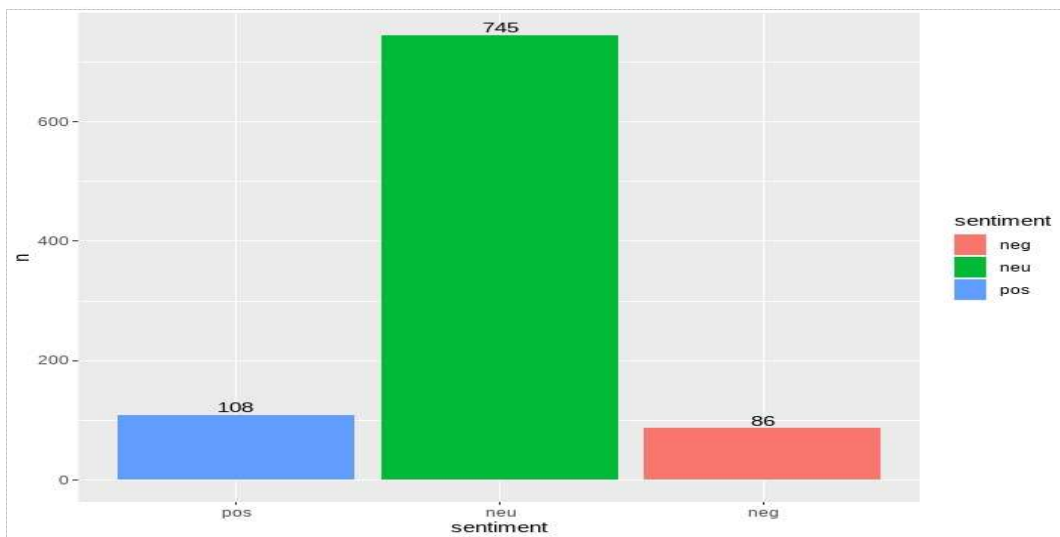
이같은 한계에도 불구하고 필자는 ‘에타’의 유학생 관련 게시글들이 어떠한 감정을 내포하고 있는지 분석하였다. 분석 도구는 R프로그램의 KoNLP패키지의 Simplepos22함수를 활용했다. 데이터 정제 과정에서 이모티콘이나 기호, 그리고 숫자만으로 작성된 댓글은 삭제하였다. 댓글의 감정 분석은 ‘KNU 한국어 감성사전’을 사용했다. KNU 한국어 감성사전²⁹⁾은 14,854개의 감정을 나타내는 단어(word)와 감정의 강도를 표현한 숫자(polarity)로 이루어져 있다. 구체적으로 살펴보면 긍정 단어

27) 위의 글, 220쪽.

28) 강진구, 「텍스트마이닝(Text Mining)을 통해 본 제주 예멘 난민-네이버 뉴스 댓글을 중심으로」, 『다문화콘텐츠연구』 30, 중앙대학교 문화콘텐츠기술연구원, 2019, 105쪽.

29) KNU 한국어 감성사전은 <https://github.com/park1200656/KnuSentiLex>(2021.6.25. 검색)를 통해 다운받아 사용할 수 있는데, 이 감성사전이 한글 텍스트의 감정 분석에 적합한가에 대해서는 다양한 이슈가 있다.

4,871개, 부정 단어 9829개, 중성 단어 154개로 구성되어 있다. 필자는 댓글에 등장한 단어에 감성 사전의 감정 점수를 부여한 다음 점수를 합산하여 개별 댓글의 감정을 파악하는 방식을 택했다. 수집된 게시글은 ‘KNU 한국어 감성사전’에 근거해 감정 점수를 부여하였다. 감정 점수가 1 이상이면 긍정(pos), -1 이하면 부정(neg), 그 이외에는 중립(neu)으로 분류하였다. 결과는 <그림 7>과 같다.



<그림 7> 감정 점수에 따른 ‘에타’ 게시글 분류

<그림 7>은 감정 점수에 따라 ‘에타’ 게시글을 그래프로 시각화한 것이다. 분석 대상 게시글은 중립이 79.3%로 다수를 차지했고, 긍정이 11.5%로 부정이 9.1%보다 약간 많았다. 긍정과 부정에 자주 사용된 어휘를 추출하면 게시글의 내용과 함께 감정 또한 파악할 수 있다.

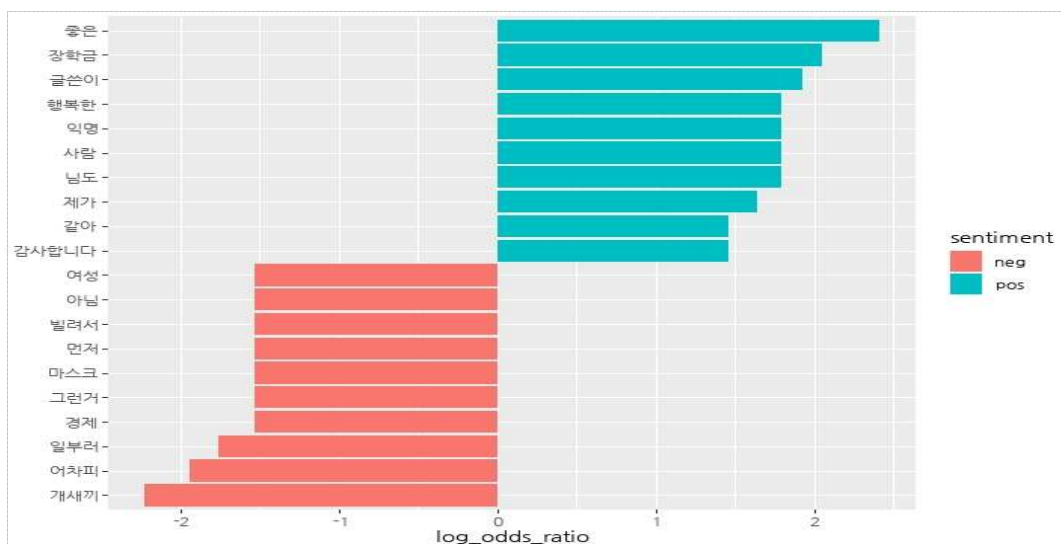
<표 3> ‘에타’ 게시글의 긍·부정 단어 빈도 상위 10

순위	단어	빈도수	감정	순위	단어	빈도수	감정
1	장학금	17	긍정	1	진짜	20	부정
2	너무	13	긍정	2	나도	10	부정
3	좋은	12	긍정	3	중국	9	부정
4	많이	9	긍정	4	그래도	8	부정

5	있는	8	긍정	5	지금	8	부정
6	진짜	8	긍정	6	개새끼	7	부정
7	같은데	7	긍정	7	그냥	6	부정
8	근데	7	긍정	8	그래서	6	부정
9	글쓴이	7	긍정	9	시진핑	6	부정
10	물론	7	긍정	10	그렇게	5	부정

<표 3>은 ‘에타’ 게시글을 긍정과 부정으로 분류한 후, 어떤 단어가 자주 사용되고 있는지를 추출한 것이다. 이를 통해 긍정과 부정의 대략적인 차이를 파악할 수 있지만, ‘진짜’ 같은 단어는 긍·부정 모두에서 높은 빈도를 보인다는 점 또한 확인 된다. 따라서 긍정과 부정의 차이를 보다 명확하게 이해하려면 양쪽 감정어에서 상대적으로 자주 사용하는 단어를 비교해야 한다.

필자는 로그 오즈비 값이 0보다 크면 긍정으로, 0보다 작으면 부정으로 분류하였고, 로그 오즈비 값이 가장 큰 상위 단어 10개를 추출하여 그래프화하였다.



<그림 8> ‘에타’ 게시글의 상대적 감정어 빈도수 상위 10

① 너무 귀엽다 저도 이제 졸업이라 와닿네요TT 저에게도 제주는

참 행복한 기억..! 사회에 나가서도 화이팅해요!!♡ 익명11님한테도 행복한 기억이라 다행이네요! 앞으로도 행복한 일 가득하길 바라요. 화이팅!!♡ 8

②재밌는 추억 남기셨네요! 앞으로도 행복한 일만 있을거예요 어디서든 화이팅! ㄴ소중한 추억 잘 간직할게요! 익명13님도 일이 원하는 대로 잘 되셨으면 좋겠어요 8

③선생님. 제 답변은 피하시고 원론적으로 돌아가려 하시는데, 프레임임을 벗어나는 방법중의 하나지요 인정합니다. 조선족이 저질렀다 생각 안합니다. 위에 본문의 요지는, 해 현상의 주체는 공산당의 지령을 받은 조선족이 그러할 가능성이 크다 라는 겁니다. 선생님, 선생님은 이 일에 대하여 제 의견에 대한 비판만을 하시지 선생님의 의견은 매우 모호하거나 언급을 하지 않으시네요. 부탁드립니다. 선생님의 의견을 말씀 해 주세요.-4

④님이 비꼬는 것만 보이지 글쓴이가 님을 비꼰건 안 보이는데요? 나도 유학생생활해봤고 내 친구들도 캐나다 미국 여러 곳에서 유학을 하고 있지만 이런 식으로 유학생 비꼬는 거 본인한테 얼마나 상처인 줄 알아요? 익명의 힘을 빌려서 유학생이라서 때리고 싶다고 말하는게 존대예요? -4(숫자는 감정 점수)

인용문은 긍정과 부정으로 판별된 글 중 상위에 랭크된 글 2개를 무작위로 선택한 것이다. ‘KNU 한국어 감성사전’에 의한 감정 분석은 누적 점수의 ‘+’ 와 ‘-’ 값으로 댓글의 긍·부정 여부를 판별한다. 또한, 점수의 총합으로 감정의 강도를 판정하는 알고리즘을 택하고 있다. 따라서 이 분석법은 사용된 단어의 개수가 모두 상이한 글을 대상으로 감정 점수를 부여하고, 그렇게 부여된 감정 점수 값을 기준으로 감정의 강도를 판별할 수 있는가 하는 이슈가 남는다.³⁰⁾ 하지만 이 같은 이슈에도 불구하고 필자는 분석을 시도하였는데, 그 이유는 댓글 감정 점수를 통해 댓글의 전반적인 감정 상태를 파악하기가 편리하기 때문이다.

30) 김영우, 앞의 책, 226쪽.

긍정의 최고 점수가 8임에 비해 부정 글의 최고 점수가 -4을 기록한 것으로 보아 외국인 유학생 관련 ‘에타’의 게시글이 부정댓글보다 긍정의 강도가 높게 나타나고 있음을 확인할 수 있다.³¹⁾

‘KNU 한국어 감성사전’에 의한 감정 분석에서 감정 점수가 +가 되어 긍정으로 분류된 글들도 실제 읽어보면 매우 부정적인 감점이 표출되고 있음을 확인할 수 있다. 가령 긍정글 중에서 로그 오즈비 값이 가장 큰 단어로 분류된 ‘좋은’이란 단어가 사용된 글을 통해서 확인해 보도록 하자.

- ① 다른 사람들은 다 예의있게 반박하는데 가만보니까 익 애는 한결 같이 비비꼬아 말하고 욕박네 반박하는건 **좋은데** 익명성 뒤에 숨어서 욕은 하지 말렴 질 떨어지누
- ② 글쓴분이 댓글에 달린 의견도 다 이해하셨다고 적었는데 밑에 비하 댓글은 왜 달아 개념 좀 챙김시다 그리고 갑자기 기준이 올라 당황스러운 건 이해하지만 한국인 학생들보다 장학금 받기 더 **좋은** 조건인 건 맞는 것 같아요 특수한 상황이기도 하고 절대평가인 만큼 더 널널하게 주시는 교수님도 있으시니 힘내세요
- ③ 욕하고 싶은 친구 있음 쪽지 보내도 **좋은데** 나는 다른 친구들의 견도 듣고싶은거라 댓글이 더 좋아
- ④ 사람을 싫어하지 국적을 싫어하진 마 그니까 국가는 싫어해도 될지라도 그 국적때문에 사람 싫어하진 말라는 뜻 병신새끼 무비자로 건너온 짱개 감염자한테 바이러스 옮겨서 숨 못쉬다 뒤질때도 그 소리 나오나 보자 우한으로 의료봉사라도 가지 그러냐 **좋은** 사람도 만에 하나 잇겠다는 취지로 말한듯 너무 비방말자 (**강조는**

31) 이 부분은 좀더 세밀한 논의가 필요하다. 일반적으로 익명의 게시글이나 댓글의 경우 부정의 감정이 훨씬 크지만, 본 연구에서 수집된 자료의 분석은 부정보다 긍정이 더 높게 나온 것으로 판별되었다. 그 원인 다양하게 추론할 수 있지만, 우선 수집된 데이터의 양이 비교적 적고, 최근 ‘에타’의 혐오표현 등이 사회적 이슈가 되면서 다수의 게시글이 관리자에 의해서 삭제되었기 때문으로 판단된다.

필자가)

인용문은 긍정으로 분류된 글로써 <그림 8>에서 보듯이 ‘로그 오즈 비’ 값이 가장 큰 ‘좋은’이란 단어가 사용된 글 중 일부이다. 글에는 ‘좋은’이란 긍정적인 어휘가 사용되고 있음에도 상대방에 대한 혐오와 분노, 심지어는 조롱의 감정이 직설적으로 표출되어 있다. ④의 인용문에서 확인할 수 있듯이 ‘좋은’의 긍정어는 ‘병신새끼’, ‘짱개’, ‘감염자’, ‘바이러스’, ‘뒤질 때’, ‘우한’, ‘비방’ 등 비판 대상을 비하하는 단어와 결합됨으로써 중국인 유학생이나 유학생에 대해 열린 태도를 견지해야 한다는 주장에 대한 강력한 혐오의 감정을 표출한다.

따라서 감정 점수의 총합으로 감정 상태를 판별하는 현재의 분석틀로는 같은 단어라도 문맥에 따라 그 감정과 의미가 달라지는 한국어의 미묘한 감정 표현을 정확하게 분석하고 해석하기에는 일정한 한계를 보인다.

4. 의미망 분석

감정 분석을 통해 우리는 한국어는 문맥에 따라 그 감정과 의미가 달라진다는 사실을 확인하였다. 따라서 유학생에 대한 정확한 감정을 이해하기 위해서는 댓글을 구성하는 단어들이 어떤 맥락에서 사용되었는지를 파악하는 것이 필요하다. 이를 위한 분석 방법이 의미망 분석(SNA: Semantic Network Analysis)이다. 의미망 분석은 단어들이 어떤 단어와 관련 맺고, 의미 형성에 중요한 역할을 하는지를 가시적으로 보여준다.

필자는 ‘에타’ 게시글의 단어 간의 관계를 살펴보기 위해 R프로그램의 widyr패키지를 이용했다. widyr패키지의 “pairwise_count”라이브러리는 동시 출현 단어 분석(co-occurrence analysis)을 제공해 준다. 결과는 <표 4>와 같다.

〈표 4〉 ‘에브리타임’ 게시물 동시 출현 단어 빈도 상위 10

순위	단어1	단어2	빈도수	순위	단어1	단어2	빈도수
1	없다	아니다	21	11	문제	아니다	16
2	아니다	같다	20	12	외국	유학생	16
3	없다	같다	18	13	유학생	아니다	15
4	유학생	같다	18	14	외국	같다	14
5	중국인	유학생	18	15	좋다	같다	14
6	생각	같다	17	16	사람	같다	13
7	중국	아니다	17	17	중국	그렇다	13
8	그렇다	아니다	17	18	많다	아니다	13
9	그렇다	같다	16	19	정도	아니다	13
10	없다	그렇다	16	20	유학생	많다	13

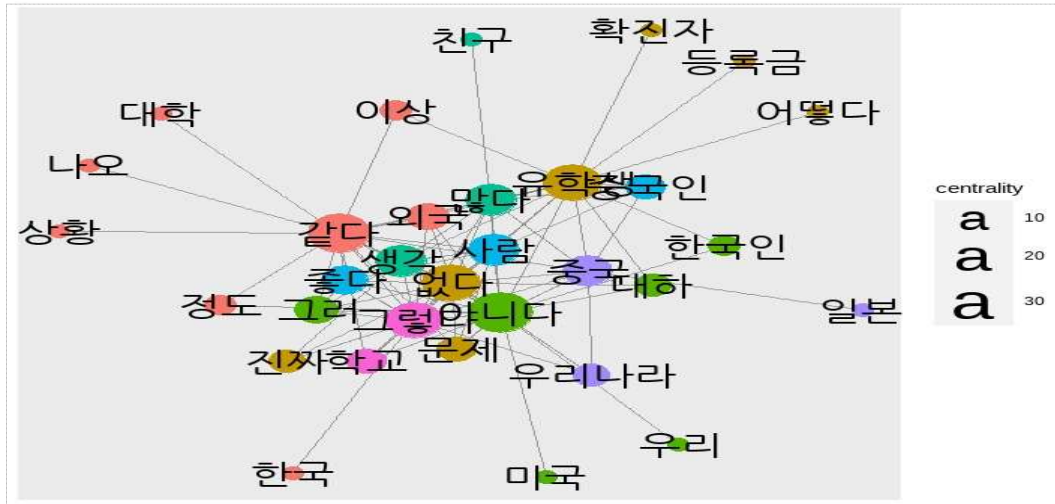
<표 4>를 보면 ‘없다’와 ‘아니다’가 21번 함께 사용되어, 전체 게시물에서 가장 많이 사용된 단어쌍임을 볼 수 있다. 그 다음이 20회의 출현 빈도를 기록한 ‘아니다’와 ‘같다’라는 단어쌍이다. 그런데 “pairwise_count” 라이브러리는 한 단어를 기준으로 함께 출현하는 단어의 빈도를 구하기 때문에 프로그램상으로는 ‘없다’-‘아니다’와 ‘아니다’-‘없다’가 같은 빈도로 출현한다. 필자는 단어1과 단어2의 역 단어쌍도 같은 의미를 갖기에 <표 4>에서 생략하였다.

위 분석을 통해 우리는 특정 단어와 자주 함께 사용된 단어를 추출할 수 있다. 다시말해, 특정 어휘가 어떤 어휘와 결합해 의미를 형성하는지를 파악하게 해준다. 예를 들면 ‘입국금지’라는 단어는 ‘그렇다’, ‘중국’, ‘사람’, ‘일본’, ‘아니다’, ‘중국인’ 등의 단어와 함께 등장하면서 일정한 의미를 형성한다.³²⁾ 동시 출현 단어들은 네트워크 분석을 통해 시각화하면 특정 단어가 어떤 맥락에서 사용되고 있는지를 한 눈에 확인할 수 있다.

필자는 R 프로그램의 “ggraph” 라이브러리를 사용하여 그래프로 나타

32) 참고로 ‘베트남’이란 단어는 ‘중국’, ‘사람’, ‘없다’, ‘진짜’, ‘새끼들’, ‘유학생’ 등의 단어와 함께 추출된다.

냈는데, 결과는 <그림 9>과 같다.



<그림 9> 동시출현 단어 네트워크 분석

필자는 ‘에타’의 게시글을 대상으로 동시 출현 빈도를 조사하여 동시 출현 빈도 8이상 인 단어에 대해 네트워크를 분석하였다. <그림 9>는 네트워크 분석 그래프를 바탕으로 의미 파악을 보다 편리하게 하기 위해 연결 중심성(degree centrality)과 커뮤니티(community)를 추가한 것이다. 연결 중심성이 높으면 글자와 원의 크기가 상대적으로 크며, 동일한 커뮤니티 별로 색깔이 구별되어 있기에 네트워크의 구조를 쉽게 이해할 수 있다.

네트워크 그래프의 이 같은 특성을 염두에 두고 <그림 9>를 해석하면, 사람을 중심으로 ‘많다’, ‘중국’, ‘없다’, ‘생각’, ‘같다’, ‘유학생’ 등의 어휘가 강력한 연결 중심성을 형성함으로써 댓글 전체의 내용을 지배하고 있다. 또한, 네트워크의 중심에 위치 한 ‘사람은 ‘좋다’, ‘중국인’ 등의 어휘와 같은 커뮤니티를 형성함으로써 ‘유학생’이나 ‘아니다’와는 다른 의미를 형성하고 있다. 또한 ‘유학생’은 ‘확진자’, ‘등록금’, ‘어떻다’, ‘없다’, ‘문제’, ‘진짜’ 등의 단어와 커뮤니티를 이루고 있다.



〈그림 10〉 바이그램 네트워크 분석

<그림 10>의 그래프는 ‘에타’의 게시글들이 어떻게 의미를 구성하는지를 파악하기 위해 바이그램 분석을 실시한 후, 동시 출현 4회 이상 어휘를 산출해 시각화한 것이다. 바이그램 네트워크 분석은 게시글의 의미구성을 파악하는 좋은 연구 방법이지만, 단어쌍의 종류가 적고 빈도도 낮다는 특징을 갖고 있다. 따라서 동시 출현 횟수를 얼마로 설정하느냐에 따라 의미파악의 난이도가 결정되기에 신중한 선택이 요구된다.

노드가 연결되어 있는 <그림 9>와 달리 <그림 10>의 그래프는 관련성이 큰 단어들끼리만 연결되어 있어 의미 파악이 더욱 명확하게 드러나 있다. 바이그램 네트워크는 ‘에타’의 게시글이 총 16개의 의미 단위로 형성되어 있음을 보여준다. 단어 집합을 통해 주제를 도출하는 토픽모델에서는 해석상의 편의를 위해 <그림 5>에서 가장 높은 성능 지표 값으로 제시한 16을 제외했는데, 노드 간의 연결을 통해 직접적으로 글의 의미 생산 과정을 보여주는 네트워크 분석이 제외된 값에 가까운 결과를 제시하고 있다는 점에서 상호 보완적인 분석이 필요하다고 판단된다.

바이그램 네트워크를 분석해 보면, 원이 가장 큰 ‘유학생’-‘등록금’

단어쌍은 ‘중국인’, ‘외국’, ‘국적’, ‘방글라데시’와 결합되어 외국인 유학생이나 유학생의 국적 문제, 또는 중국인 유학생, 유학생의 등록금 등과 관련된 의미가 생성되고 있음을 알 수 있다. 비슷한 방법으로 장학금을 받거나, 장학금 커트라인이 얼마 이상인지에 대한 논의, 개강을 맞이한 캠퍼스의 풍경 등이 사람 많다, 사람 없다, 진짜 많다 등의 연속된 단어 구성을 통해 제시되고 있다. 이 밖에도 설문 작성이나, 중국인 입국금지, 그리고 문재인 지지자, 개강 연기 등도 게시글의 중요 내용을 형성하고 있음을 알 수 있다.

IV. 맺음말

필자는 COVID-19 팬데믹으로 인해 트랜스내셔널이 위협받는 상황에서 외국인 유학생에 대한 대학생들의 생각을 텍스트마이닝 분석을 통해 살펴보았다.

분석 대상은 ‘에브리타임’에 게시된 외국인 유학생 관련 게시글 중 중앙대와 제주대의 게시글 1,237건이며, 기간은 2019년 12월 1일부터 2021년 8월 31일까지로 한정하였다. 분석 방법으로는 빈도 분석, 토픽 분석, 감정 분석, 의미망 분석을 시도했다. 그 결과는 다음과 같다.

첫째, 중국 또는 중국인 유학생에 대한 스테레오타입을 대학생들 역시 반복, 확대, 재생산하고 있었다. 국가인권위원회 등에서 중국인에 대한 혐오표현이라고 밝힌, ‘착짱죽짱’, ‘짱깨’, ‘쓰레기’, ‘미개’, ‘냄새’, ‘공산당’, ‘중국산’, ‘좃선족’, ‘우한폐렴’, ‘차이나게이트’, ‘살인’, ‘시끄럽다’ 등의 어휘가 게시글과 댓글에서 아무런 비판의식도 없이 남용되고 있다.

둘째, 한국의 대학생들은 트랜스내셔널과 초연결성사회라는 세계사적 흐름 속에서 어느새 자신들과 대학 공간을 공유하고 있는 유학생들에 대해 동의보다는 반대를, 수용보다는 거부의 태도를 보인다. 이같은 인식으로 인해 대학생들은 유학생의 편의를 위한 대학 당국의 일련의 조

치에 대해 즉자적인 형태로 역차별의 감정을 표출하는 특징을 보이고 있다.

셋째, 코로나19 팬데믹과 관련해 일부 유학생의 부적절한 행위에 대해 대학생들은 유학생 개인의 문제가 아닌 집단의 속성으로 규정한 후, 병적이라 할 만큼 그 집단에 대한 극단적인 혐오의 감정을 표출한다. 글로 형언할 수 없을 정도의 중국인에 대한 혐오 표현 등을 보고 있으면, 무엇이 대학생들을 이렇게까지 극단적으로 내몰았는가 하는 의문이 들기도 한다. 자신보다 지위가 낮다고 생각되는 상대(소수자)에 대한 무시와 조소, 그리고 신인종주의에 근거한 우월의식의 표출은 역설적으로 오늘날 대학생이 얼마나 불안한 위치에 있는지를 보여주는 증거이기도 하다.

넷째, 유학생에 대한 대학생들의 차별 행위에는 트랜스내셔널과 초연결사회, 코로나19 팬데믹으로 인해 기존의 질서가 붕괴되었지만, 아직 그것을 대체할만한 새로운 질서가 형성되지 않는 ‘궤위(闕位: interregnum)의 시대’에 놓여 있는 한국 대학사회의 슬픈 자화상이자, 불안과 위협 속에 방치된 채 경쟁으로 내몰린 대학생들의 처절한 자기방어 기제의 흔적이 묻어 있다.

이상의 분석은 ‘궤위(闕位: interregnum)의 시대’를 넘기 위해서라도 대학 교육 과정의 변화는 필수적이며, 특히 교양 교육은 세계사적 변화를 반영할 수 있는 내용과 방법을 모색해야 할 필요가 있다. 필자는 그러한 방안으로 ‘문화적 다양성’과 ‘지속가능한 발전’에 대한 이해와 동의, 더 나아가 수용까지를 고려한 보다 적극적인 교육프로그램의 개발과 그 실천을 제안한다. 구체적인 교육 프로그램에 대해서는 추후 논의를 통해 보완되길 기대한다.

참고문헌

1. 기초자료

네이버뉴스(<https://news.naver.com>)

에브리타임(<https://everytime.kr>)

2. 논저

「혐오 글에 유학생 빨났다」 《중대신문》, 2018.10.8.

Desvinta Ayu Iriani, Ito Hiroko, 「대구경북지역 거주 아시아권 출신 유학생의 학교 내 차별경험 출신국가지역에 따른 위계적 차별」, 『한국사회학회 사회학대회논문집』, 한국사회학회, 2018.

ZHOU NAN, 「중국과 관련된 코로나-19 뉴스이용이 중국인에 대한 제노포비아 정서에 미치는 영향」, 서강대 석사학위, 2020.

강운선, 「외국인 유학생에 대한 한국 대학생의 인식 유형」, 『현대사회와 다문화』 6(2), 대구대학교 다문화사회정책연구소, 2016.

강진구, 「빅데이터 분석을 활용한 다문화 연구 동향 분석」, 『다문화콘텐츠연구』 35, 중앙대학교 문화콘텐츠기술연구원, 2020.

_____, 「팬데믹과 혐오-베트남 관련 기사의 댓글을 중심으로」, 『다문화콘텐츠연구』 37, 중앙대학교 문화콘텐츠기술연구원, 2021.

_____, 이기성, 「텍스트마이닝(Text Mining)을 통해 본 제주 예멘 난민-네이버 뉴스 댓글을 중심으로」, 『다문화콘텐츠연구』 30, 중앙대학교 문화콘텐츠기술연구원, 2019.

교육부, 『2020년 국내 고등교육기관 내 외국인 유학생 통계』, 교육부, 2020.

권상희·김익현, 「온라인 댓글 인식과 댓글 활동의 관계에 관한 연구: 댓글의 신뢰도와 인터넷 뉴스 수용자의 수용 경향을 중심으로」, 『한국언론정보학보』 42(2), 한국언론정보학회, 2008.

김세경, 「외국인 유학생의 문화적응스트레스가 대학생활적응에 미치는 영향」, Journal of the Korean Data Analysis Society 19(2), 2017.

김영우, 『Do it! 쉽게 배우는 R 텍스트 마이닝』, 이지스퍼블리싱, 2021.

- 김정숙, 「초국적 주체와 COVID-19: 유학생의 인종언어 이데올로기 차별 경험, 정체성, 그리고 행위주체성」, 『현대사회와다문화』 11(2), 대구대학교 다문화사회정책연구소, 2021.
- 김정희·김영순, 「재한 중국동포 유학생의 문화적응 과정을 통해 본 정체성 협상과 그 의미」, 『겨레어문학』 60, 겨레어문학회, 2018.
- 김종태·한기덕, 「한국 대학생의 외국인 차별 의식의 근원-민족 문명 선진국 담론의 비판적 검증」, 『담론201』 16(3), 한국사회역사학회, 2013.
- 김해연·신리나·황하성, 「중국인 유학생의 차별경험이 유학생생활 만족도에 미치는 영향 귀인성향, 내외통제성의 조절효과」, 『사회과학연구』 26(1), 동국대학교 사회과학연구원, 2019.
- 문화체육관광부 해외문화홍보원, 『2020년 국가이미지 조사 보고서』, 문화체육관광부 해외문화홍보원, 2021.
- 민예슬·김창대, 「외국인 유학생의 일상적 차별경험이 매일의 우울에 미치는 영향: 내부귀인의 조절효과」, 『상담학연구』 19(6), 한국상담학회, 2018.
- 백근영·김애란, 「재한 중국 여자 유학생의 대인관계에서의 성적 대상화 경험이 섭식행동에 미치는 영향: 지각된 차별감의 조절된 매개효과」, 『한국심리학회지: 상담 및 심리치료』 32(4), 한국심리학회, 2020.
- 백영민, 『R를 이용한 텍스트 마이닝』 개정판, 한울, 2020.
- 이예지, 「대학생의 필수 앱 ‘에브리타임’의 명과 암」, 《소비라이프》, 2021.
- 임운서·안운정, 「대학생의 잠재된 차별의식 드러내기-차별대상과 편견 성찰에 대한 질적 연구」, 『민주주의와 인권』 16(4), 전남대학교 5·18연구소, 2016.
- 임희진·김제희·정소현·정하영·김현성, 「캠퍼스 국제화에 대한 국내 학생들의 인식 연구」, 『아시아교육연구』 16(1), 서울대학교 교육연구소, 2015.
- 중앙대학교 학생생활상담연구소, 「2017 유학생 의식조사 및 정신건강실태 보고」, 중앙대학교학생생활상담연구소, 2017.

- 진사선, 「캠퍼스내 중국인 유학생들의 차별경험-개념도방법을 중심으로」, 중앙대 석사학위논문, 2019.
- 출입국·외국인정책본부, 『출입국·외국인정책 통계월보 9월호』, 출입국·외국인정책본부, 2021.
- 황동진·김혜영, 「신인종주의적 관점으로 분석한 유학생 생활만족도에 영향을 미치는 요인」, 『한국콘텐츠학회논문지』 21(5), 한국콘텐츠학회, 2021.

Abstract

A Study on self-portrait of COVID-19 Pandemic and Hyperconnected Society

– Focusing on discrimination against international students
in universities –

Kang, Jin-gu^{*}

The purpose of this study is to examine Korean college students who are in an unprecedented situation due to the COVID-19 pandemic. To this end, I analyzed college students' thoughts on foreign students using text mining techniques under the COVID-19 pandemic where transnationals feel threatened. The main target of analysis was 1,237 posts from Chung-Ang University and Jeju National University among posts related to foreign students posted at Everytime, and the period was limited from December 1, 2019 to August 31, 2021. In order to compare Koreans' sentiments toward international students, a comparative analysis was also conducted with the comments on articles related to international students on Naver News during the same period. The collected data were analyzed using packages such as “KoNLP”, “WordCloud2”, “ggraph”, and “topicmodels” of the R program, and the results are as follows.

First, Korean college students also repeated, expanded, and reproduced stereo types of Chinese or Chinese international students prevalent in Korean society, such as “Chakjjangjukjjang”, “Jjangkae”, “Jotseonjo”, “China Gate”, and “Wuhan pneumonia”.

^{*} Research Professor, Cumda Humanities JEJUNU

Second, Korean college students showed an attitude of disapproval rather than consent and rejection rather than acceptance of the existence of international students who share space with them or changed university system amid the global trend of transnational and hyperconnected society.

Third, Korean college students defined the inappropriate behavior of some international students related to COVID-19 as an attribute of the group, not an individual problem, and expressed extreme hatred toward the group.

Fourth, Korean college students' discrimination against foreign students requires special attention in that it is a self-defense of them who have endured the 'era of interregum' while left unattended in anxiety and threats.

Key-words: Text mining, College students, International students, Evertime, Discrimination, Hate speech, COVID-19, Pandemic

논문투고일 2022. 2. 13.

심사완료일 2022. 2. 28.

게재확정일 2022. 3. 15.