

# 저소득층 대학생들의 진로준비과정에서의 성별·전공별 특성에 대한 사례연구: 텍스트 빈도분석과 연관분석의 적용

이지혜<sup>1</sup>, 이신혜<sup>2\*</sup>

<sup>1</sup>한림대학교 교양기초교육대학 교수, <sup>2</sup>서울대학교 교육학과 박사과정

## A Case Study on Characteristics of Gender and Major in Career Preparation of University Students from Low-income Families: Application of Text Frequency Analysis and Association Rules

Jihye Lee<sup>1</sup>, Shinhye Lee<sup>2\*</sup>

<sup>1</sup>Professor, College of General Education, Hallym University

<sup>2</sup>PhD student, Department of Education, Seoul National University

요 약 이 연구는 청년들의 높은 비정규직 비율과 계층 양극화의 우려 속에서 저소득층 대학생의 진로준비과정에 대해 이해하고 시사점을 구하기 위한 것이다. 이를 위하여 S 장학재단에서 장학금 지원을 받는 13명의 대학생들을 연구 대상으로 선정하였고, 6회의 인터뷰를 진행하여 그 축어록을 바탕으로 텍스트마이닝 기법을 활용한 분석을 실시하였다. 분석 결과, 대학생들은 인터뷰 과정에서 이전의 학업 경험을 회상하거나 진로를 설계할 때, 가정환경과 소득수준의 영향을 받는 것으로 보이며, 이러한 차이는 성별, 전공별로 다른 특성이 있는 것으로 나타났다. 이 연구는 질적 연구 방법으로 축적된 자료에 텍스트마이닝 기법을 융합적으로 적용하여 분석한 연구로 종래의 진로연구에 비하여 방법론적 확장을 시도했다는 의의를 갖고 있다. 그 결과, 저소득층 대학교 장학생들의 성별 및 전공별 진로준비과정의 차이를 대학생활 및 진로준비와 관련된 단어들의 관계를 통해 탐색적으로 살펴볼 수 있었다.

주제어 : 대학생, 저소득층, 진로발달, 장학생, 진로준비, 연관분석

**Abstract** This study aims to understand and to infer the implications from the career preparation experiences of low-income university students in the context of high youth unemployment rate and the polarization of the social classes. For this purpose, we selected 13 university students who received scholarship from the S scholarship foundation and conducted analysis using text mining techniques based on the six-time interviews. According to the results, university students seem to be influenced by home environment and income level when recalling previous academic experience or designing career during the interview process. Also, these differences were found to have different characteristics according to gender and major. This study is meaningful in that the qualitative research data is analyzed by applying the text mining technique in a convergent way. As a result, the college life and career preparation of low-income university students were explored through the frequency and relation of words.

**Key Words** : University student, Low-income class, Career development, Scholarship, Career preparation, Association rules

\*This research was supported by Hallym University Research Fund, 2017(HRF-201711-016).

\*Corresponding Author : Shinhye Lee(lee.sinhye@snu.ac.kr)

Received October 15, 2018

Revised November 26, 2018

Accepted December 20, 2018

Published December 28, 2018

## 1. 서론

우리 사회에서 계층간 교육격차의 문제는 지속적으로 제기되어 왔으나 대학단계에서의 계층 간 경험 차이에 대한 주목은 비교적 최근의 일이다. 즉, 부모의 사회·경제적 지위가 대학진학 이후의 삶과 경험에도 지속적인 영향을 미치고 있다는 것이다[1-6]. 급속한 고학력자의 팽창에 따라 노동시장에서 학력 자체가 지니는 사회계층의 분화 기능이 점점 약화되고 있는 반면, 동일(혹은 유사)한 학력 내에서도 학교의 명성, 차별화된 교육적 경험 등을 통한 학력의 질적 분화는 점점 강화되고 있어서, 고등교육에서의 질적 분화는 고등교육의 진입단계뿐만 아니라 진입이후에서도 발생하고 있으며, 양자 모두에서 학생들이 갖는 사회적 계층에 따른 불평등이 존재한다[3]. 박경호 등의 연구에서도, 소득수준으로 측정된 부모의 사회·경제적 지위와 대학유형은 대학 졸업생의 첫 일자리의 노동시장 지위를 결정하는 주요 요인일 뿐만 아니라, 어학연수, 휴학, 취업프로그램, 정부정책 등의 대학에서의 경험 역시 대학 졸업생들의 노동시장 지위에 영향을 미치는 주요 요인임을 확인한 바 있다[6].

위와 같은 선행연구는 대학교육에서의 학생들의 계층간 경험의 차이가 상존하며, 그 영향이 대학생활과 졸업 이후의 소득지위 등에 영향을 미친다는 것을 보여준다. 실제로 여러 경험적 연구에 따르면, 저소득층 대학생은 학비와 생활비 마련의 과제로 중산층 이상의 대학생과 달리 여러 경제적 어려움에 시달리고, 향후의 진로와 관련해 여러 측면에서 취약한 상태에 머문다[7-10]. 이지혜·황매향은 저소득층 대학생은 전공 트랙에 진입하면서 본격적인 진로준비를 시작하지만, 부담이 되는 가족 상황과 부족한 진로 자원으로 인하여 애로를 겪으면서, 현실에 맞는 진로포부의 '조정', 최소의 자원을 효율적으로 활용하기 위한 '관리', 그리고 네트워크의 '확장'을 통해 자신의 진로준비 상황을 개선하려는 시도를 한다고 묘사하고 있다[11].

이러한 연구들은 계층간 경험의 차이와 저소득층의 애로를 제기하고 있으나, 저소득층 대학생 내 경험과 인식의 차이에 대해서는 드러내지 못하고 있다. 이에 본 연구에서는 우수한 저소득층 대학생을 선발하여 지원하는 S장학재단의 장학생을 대상으로, 질적 종단연구과정에서 수집된 자료를 바탕으로 하여 저소득층 대학생의 진로경험이 성별과 전공계열에 따라 어떠한 내적 차이가 있는지를 분석하고자 하였다. 이를 위해 텍스트마이닝

기법인 텍스트 빈도분석과 연관분석을 활용하였는데, 이는 비정형 데이터인 텍스트에서 의미 있는 정보를 추출하기 위한 분석방법이다. 텍스트마이닝 기법을 이용하여 추출된 단어들은 하나 또는 그 이상이 결합하여 개념(concept)을 형성하며, 두 개념 간의 연계는 서술(statement)이 되고, 이는 언어지도(semantic map)를 형성한다고 알려져 있다[12]. 본 연구에서는 이러한 단어들의 빈도와 단어 간 관계를 양적으로 분석하고 이를 시각화하여 집단 별로 차이가 나타나는지 분석하고자 하였으며, 이를 통해 도출된 연구 결과와 연구 방법을 추후 연구를 위한 기초자료로 탐색적으로 활용해 보고자 하였다.

## 2. 연구 방법

### 2.1 연구 대상

본 연구의 대상은 2014년부터 2016년까지 대학교에 재학(휴학 포함) 중이었던 13명의 대학생들로, 2014년 표집 당시 대학 2학년에 재학중이었던 학생들이다. 또한 이 학생들은 재단활동참여에 대한 성실성 기준과 최저학점 기준을 만족한 장학생들로, 가족구조, 경제상황, 성별 등을 고려하여 표집선정되었다. 따라서 이들은 저소득층 대학생 일반이라기 보다는 저소득층 가정배경을 가진 장학생이라는 특수한 사례집단이다. 이 학생들은 대학2학년 1학기 이후, 매 학기마다 진로와 관련한 6번의 인터뷰에 참여하였으며, 본 연구에는 참조인과 함께 면담한 일부 회기를 제외하고 5번의 인터뷰 자료(2014년 2회, 2015년 2회, 2016년 1회)를 분석에 활용하였다. 13명의 학생 중 8명(61.5%)은 여학생, 5명(38.5%)은 남학생이었고, 전공에 따라서는 인문계열(사회·경상계열 포함) 5명(38.5%), 이공계열 5명(38.5%), 교육계열 3명(23.1%)이었다. 분석을 위해 사용된 인터뷰 자료는 13명의 5회기 인터뷰 자료로, 약 65시간에 해당하는 축어록 내용을 분석하였다.

### 2.2 분석 방법

본 연구는 대학생들의 진로 발달 과정과 이에 영향을 미치는 요인에 대해 탐색적으로 분석하고자 텍스트 빈도 분석과 연관성 분석을 활용하였다. 이를 위해 R 프로그램(버전 3.5.0)을 활용하였으며, 자연어 처리와 데이터 정제, 연관분석, 시각화를 위해 tm, KoNLP, stringr, tdm, wordcloud2, arules, qgraph, arulesViz 패키지를 사용하였다.

### 2.2.1 빈도 분석

단어 빈도는 전체 자료 및 성별, 전공별 특성에 따라 분류된 자료에서 얼마나 특정한 단어가 자주 등장하는지를 나타내는 지표로, 빈도가 높을수록 해당 단어가 대화에서 자주 등장하는 핵심적인 단어임을 의미한다. 따라서 본 연구에서는 연구절차를 통해 의미 있는 단어에 대한 빈도 분석을 시행하여, 전체 자료에서 및 성별, 전공별 특성에 따라 단어의 빈도가 어떻게 다르게 나타나는지 파악하고자 하였다.

### 2.2.2 연관 분석

연관분석(Association rule mining) 기법은 데이터에서 변수들 간의 관계를 분석하기 위한 방법으로 두 개 이상의 단어가 공통으로 출현하는 빈도를 계산하여 유용한 정보나 패턴을 추출하기 위한 방법이다. 이러한 연관 규칙은 'A⇒B' 형태로 나타내는데, A는 조건부(antecedent), B는 결론부(consequent)를 의미하고, 'A⇒B'는 조건 A의 단어가 등장한 문장(문맥)에 결론인 B 단어가 등장한다는 의미이다. 연관분석에서 활용되는 지표로는 지지도(support), 신뢰도(confidence), 향상도(lift)가 있다. 지지도(support)는 조건부와 결론부에 등장하는 단어가 전체 문서에서 나타나는 빈도의 비율을 의미하고, 신뢰도(confidence)는 지지도를 조건부의 단어 등장 비율로 나눈 값으로 신뢰도가 높을수록 강한 연관 규칙임을 의미한다. 향상도(lift)는 신뢰도를 결론부의 단어 등장 비율로 나누어 규칙의 유용성을 실질적으로 판단할 수 있는 기준을 제시하는데, 향상도 값이 1보다 크면 두 단어의 관계가 독립일 때보다 유용하다고 판단할 수 있다 [13,14]. 연관분석을 위해 모든 조합의 단어들을 분석하는 것은 시간이 오래 걸리기 때문에 최소한의 지지도를 기준으로 빈발 항목집합(frequent item sets)를 재귀적으로 계산하는 Apriori 알고리즘을 이용하여 계산할 수 있다 [15].

### 2.3 분석 절차

텍스트 자료의 분석을 위해 전사된 인터뷰 내용을 반복적으로 읽고, 분석에 포함될 인사말을 제외한 인터뷰 대상자의 발화를 따로 분리하였다. 이후, '그래서', '같은', '생각' 등과 같이 본 연구의 주제와 관련이 낮은 단어와 개인 정보, 숫자, 접속사 등을 모두 삭제하여 불용어를 정리하고, 띄어쓰기, 오기를 바로잡았다. 또한 단어들 간의

관계를 파악하기 위해 2음절 이상의 단어들을 추출하였는데, 이때 단수, 복수 표현에 따라 서로 다른 단어로 추출되는 것을 방지하고 1음절의 단어가 누락되는 것을 피하고자 단수형의 단어들은 모두 '-들'을 붙여 복수형으로 바꾸어 같은 단어가 하나의 형태를 나타내도록 변경하였다. 동사만으로 문맥이 파악되지 않는 '했어요', '갔어요' 등의 동사는 삭제하였고, 이외의 단어들은 대학 생활과 직접적인 관련이 없더라도 가능한 그대로 두고자 하였다. 이러한 단어 정리와 빈도 분석 과정을 여러 번 반복하고 추출된 단어들을 연구자들이 교차로 확인하여 자료 분석을 위한 전처리를 하였는데, 이 과정에서 한 문장의 길이가 사례마다 차이가 크고 연관분석을 할 만큼 여러 단어의 공출현빈도가 크지 않아서 몇 묶음씩 묶어 분석을 진행하였다. 문맥 안에서 단어 간 관계를 파악할 수 있도록 하되, 그 길이가 적당하도록 여러 번 분석을 실시한 결과 열 문장씩을 한 단락으로 구성하는 것이 분석 가능한 최소단위라고 판단하여, 내담자 발화를 열 문장씩 한 단락으로 묶어 분석을 진행하였다. 빈도 분석을 바탕으로 추출된 단어 중 연구대상이 13명임을 고려하여 평균 1회 이상 나타난 상위 103개의 단어를 핵심단어로 선정하였으며, 이 단어들을 기준으로 단어 공출현 행렬(co-occurrence matrix)을 만들어 연관분석을 진행하였다. 본 연구에서는 이러한 방식으로 분석된 단어들과 그 결과를 바탕으로, 저소득층 대학생들의 대학생 및 진로와 관련된 단어 빈도와 단어 간 관계를 학년별(회기), 성별, 전공별로 분석하고 비교하고자 하였다.

## 3. 연구 결과

### 3.1 전체 및 회기별 빈도 분석 결과

추출된 상위 103개의 단어는 다음 Table 1과 같다. 이후 빈도 분석 및 연관분석에서는 이 103개의 단어를 기준으로 전체 비율을 산출하였다. 단어의 빈도를 고려하여 워드클라우드로 나타낸 그림은 Fig. 1과 같다. 분석 결과 인터뷰 과정에서 고등학교, 선생님, 담임선생님, 중학교, 초등학교, 가정환경 등 대학교 이전의 학업 경험과 환경을 회상하거나 비교하는 단어들의 빈도가 높았으며, 교수님, 대학교, 동아리, 선배들, 교환학생, 동기들, 멘토링 등 대학 생활과 관련된 단어가 많이 등장하는 것으로 나타났다. 긍정적, 현실적, 어려움, 심리적, 안정적 등 본인

이 느끼는 감정이나 상태를 나타내는 단어들도 일부 명사 형태로 추출된 것을 확인할 수 있다.

Table 1. List of 103 words (numbers in parentheses indicate frequency)

|  |
|--|
| 고등학교(352), 선생님(313), 친구들(265), 교수님(224), 대학교(163), 중학교(148), 대학원(137), 스트레스(134), 기숙사(126), 아르바이트(124), 남자친구(114), 장학금(109), 부모님(94), 동아리(93), 선배들(92), 재밌는(90), 실험실(89), 초등학교(87), 할머니(81), 담임선생님(79), 교환학생(69), 분위기(69), 자격증(68), 방학(65), 동기들(60), 멘토링(59), 대학생(53), 동생들(53), 프로그램(49), 학생들(48), 로스쿨(47), 여자친구(47), 자신감(46), 자기소개(45), 공무원(43), 봉사활동(42), 대외활동(41), 등록금(40), 장학생(39), 기말고사(38), 해외봉사(38), 가정환경(37), 도서관(36), 중간고사(36), 교육학(33), 긍정적(31), 생활비(28), 스터디(27), 이미지(27), 경제적(26), 룸메이트(26), 중학생(26), 핸드폰(26), 임용고시(25), 현실적(25), 고등학생(24), 프로젝트(24), 학교생활(24), 할아버지(24), 다양한(23), 적극적(22), 시험기간(21), 세미나(20), 중고등학교(20), 복수전공(19), 과사무실(19), 후배들(19), 공기업(18), 사회학(18), 수강신청(18), 영어공부(18), 책임감(18), 한국사(18), 가능성(17), 대학생활(17), 마케팅(17), 신입생(17), 어려움(17), 언니들(17), 우리나라(17), 장학재단(17), 졸업논문(17), 캠퍼스(17), 마인드(16), 모의고사(16), 스포츠(16), 심리적(16), 인간관계(16), 리더십(15), 사회생활(15), 안정적(15), 열등감(15), 초등학생(15), 커리큘럼(15), 대기업(14), 재수강(14), 배움터(13), 사회적(13), 자존감(13), 충중고(13), 취업준비(13), 페이스북(13), 학생회(13) |
|--|



Fig. 1. Word cloud for the top 103 words

회기에 따라 단어 빈도를 분석한 결과는 Table 2와 같다. 1, 2회기에는 고등학교, 친구들, 중학교 등의 이전의 경험과 관련된 단어들이 상대적으로 많이 등장하였고, 2, 3, 4 회기에는 교수님, 기숙사, 아르바이트, 선배들, 실험실, 공무원, 봉사활동 등 대학 생활 또는 진로와 관련된 단어들이 상대적으로 높은 비율을 차지하였다. 1년의 시간차가 있었던 다섯 번째 자료에서는 이전 회기에 비해 뚜렷한 특징이 없었다. 회기와 상관없이 비슷한 비율을 유지한 단어들은 스트레스, 분위기, 대학생, 동생들 등인 것으로 나타났다.

Table 2. Percentage of words per session (%)

| rate | words | session |      |      |      |      |
|------|-------|---------|------|------|------|------|
|      |       | 1       | 2    | 3    | 4    | 5    |
| 1    | 고등학교  | 10.85   | 7.96 | 2.87 | 2.92 | 7.06 |
| 2    | 선생님   | 6.49    | 4.56 | 3.41 | 4.06 | 9.27 |
| 3    | 친구들   | 6.57    | 5.63 | 2.87 | 3.62 | 5.52 |
| 4    | 교수님   | 2.88    | 4.35 | 4.30 | 7.33 | 3.02 |
| 5    | 대학교   | 4.44    | 3.18 | 2.33 | 1.33 | 3.75 |
| 6    | 중학교   | 5.67    | 2.23 | 0.18 | 0.71 | 3.61 |
| 7    | 대학원   | 1.64    | 1.91 | 3.05 | 5.57 | 1.40 |
| 8    | 스트레스  | 1.89    | 2.65 | 3.23 | 2.83 | 2.65 |
| 9    | 기숙사   | 3.45    | 3.50 | 3.23 | 1.59 | 1.10 |
| 10   | 아르바이트 | 2.38    | 3.50 | 1.79 | 2.47 | 1.77 |
| 11   | 남자친구  | 2.22    | 3.82 | 1.43 | 2.12 | 1.40 |
| 12   | 장학금   | 1.07    | 1.91 | 1.97 | 2.56 | 2.80 |
| 13   | 부모님   | 1.73    | 1.91 | 0.72 | 1.41 | 2.58 |
| 14   | 동아리   | 2.79    | 2.23 | 1.25 | 2.30 | 0.37 |
| 15   | 선배들   | 1.73    | 1.49 | 2.69 | 2.47 | 1.03 |
| 16   | 재밌는   | 2.14    | 2.02 | 2.51 | 0.97 | 1.47 |
| 17   | 실험실   | 0.33    | 1.70 | 4.30 | 2.74 | 1.03 |
| 18   | 초등학교  | 3.04    | 1.17 | 1.25 | 0.80 | 1.69 |
| 19   | 할머니   | 1.07    | 2.55 | 0.00 | 1.68 | 1.84 |
| 20   | 담임선생님 | 2.30    | 0.85 | 0.54 | 0.44 | 2.58 |
| 21   | 교환학생  | 0.74    | 1.17 | 3.94 | 1.59 | 0.66 |
| 22   | 분위기   | 1.07    | 1.49 | 1.61 | 1.41 | 1.25 |
| 23   | 자격증   | 0.90    | 0.96 | 0.18 | 3.27 | 0.74 |
| 24   | 방학    | 1.64    | 1.06 | 1.43 | 1.94 | 0.37 |
| 25   | 동기들   | 0.90    | 2.55 | 1.97 | 0.80 | 0.37 |
| 26   | 멘토링   | 0.82    | 2.44 | 1.08 | 0.80 | 0.81 |
| 27   | 대학생   | 0.90    | 1.70 | 1.61 | 0.62 | 0.74 |
| 28   | 동생들   | 0.49    | 0.96 | 1.43 | 1.41 | 1.03 |
| 29   | 프로그램  | 0.41    | 0.53 | 1.43 | 1.33 | 1.18 |
| 30   | 학생들   | 0.99    | 0.64 | 0.90 | 0.88 | 1.10 |
| 31   | 로스쿨   | 0.66    | 0.11 | 0.72 | 1.50 | 1.25 |
| 32   | 여자친구  | 1.07    | 1.27 | 0.36 | 0.80 | 0.81 |
| 33   | 자신감   | 1.07    | 0.64 | 1.08 | 0.71 | 0.96 |
| 34   | 자기소개  | 0.99    | 0.32 | 0.36 | 1.41 | 0.88 |
| 35   | 공무원   | 0.08    | 0.42 | 1.43 | 0.27 | 1.99 |
| 36   | 봉사활동  | 0.25    | 0.53 | 2.15 | 1.24 | 0.59 |
| 37   | 대외활동  | 0.99    | 0.85 | 1.08 | 0.71 | 0.52 |
| 38   | 등록금   | 0.49    | 0.53 | 0.72 | 0.80 | 1.18 |

### 3.2 성별 빈도 분석 및 연관 분석 결과

성별 연구 대상자의 수를 고려하여 분석한 단어 빈도와 비율은 다음 Table 3과 같다. 남학생들은 친구들, 부모님, 중학교, 대학교, 자격증, 등록금, 담임선생님에 대한 언급을 더 많이 하였고, 여학생들은 상대적으로 스트레스, 동생들, 교환학생, 로스쿨, 공무원에 대한 언급이 더 많은 것으로 나타났다. 이는 대학 입학 후 3년 간 남학생들은 군대를 다녀온 경우가 많았는데, 여학생들은 상대적으로 교환학생이나 대외활동을 한 경험이 있었기 때문인 것으로 보인다(Fig. 2 참고). 또한 남학생들이 전체 단어 중 ‘여자친구’를 언급한 비율(2.54%)에 비해 여학생들이 ‘남자친구’를 언급한 비율(3.25%)이 더 높았고, 남학생들은 부모님, 친구들 등 주변 지인에 대한 언급이 비교적 높은 것으로 나타났는데, 면접을 진행한 연구자는 이 차이가 상대적으로 남학생들이 기존의 대인 관계를 유지하는 경향이 있기 때문인 것으로 해석하였다.

Table 3. Percentage of words per gender (%)

| rate | words | male | female | rate | words | male | female |
|------|-------|------|--------|------|-------|------|--------|
| 1    | 고등학교  | 7.04 | 6.61   | 20   | 담임선생님 | 2.18 | 1.17   |
| 2    | 선생님   | 6.64 | 5.68   | 21   | 교환학생  | 0.45 | 1.79   |
| 3    | 친구들   | 6.37 | 4.42   | 22   | 분위기   | 0.84 | 1.58   |
| 4    | 교수님   | 4.02 | 4.45   | 23   | 자격증   | 2.07 | 0.91   |
| 5    | 대학교   | 3.91 | 2.72   | 24   | 방학    | 1.40 | 1.17   |
| 6    | 중학교   | 3.74 | 2.37   | 25   | 동기들   | 1.34 | 1.05   |
| 7    | 대학원   | 2.29 | 2.81   | 26   | 멘토링   | 1.12 | 1.14   |
| 8    | 스트레스  | 1.56 | 3.10   | 27   | 대학생   | 0.95 | 1.05   |
| 9    | 기숙사   | 1.95 | 2.66   | 28   | 동생들   | 0.11 | 1.49   |
| 10   | 아르바이트 | 2.62 | 2.25   | 29   | 프로그램  | 0.89 | 0.97   |
| 11   | 남자친구  | 0.06 | 3.31   | 30   | 학생들   | 0.50 | 1.14   |
| 12   | 장학금   | 2.23 | 2.02   | 31   | 로스쿨   | 0.17 | 1.29   |
| 13   | 부모님   | 2.79 | 1.29   | 32   | 여자친구  | 2.57 | 0.03   |
| 14   | 동아리   | 1.84 | 1.76   | 33   | 자신감   | 0.61 | 1.02   |
| 15   | 선배들   | 1.84 | 1.73   | 34   | 자기소개  | 0.73 | 0.94   |
| 16   | 재밌는   | 1.68 | 1.76   | 35   | 공무원   | 0.11 | 1.20   |
| 17   | 실험실   | 1.23 | 1.96   | 36   | 봉사활동  | 0.89 | 0.76   |
| 18   | 초등학교  | 1.90 | 1.55   | 37   | 대외활동  | 0.56 | 0.91   |
| 19   | 할머니   | 1.62 | 1.52   | 38   | 등록금   | 1.51 | 0.38   |

성별에 따른 단어 간의 관계를 살펴보기 위해 연관분석을 실행한 결과는 다음 Table 4와 같다. 분석 결과 향상도가 높은 상위 규칙에는 성별로 큰 차이가 없었으나, 상위 50개 규칙을 기준으로 단어 간 관계를 살펴본 Fig. 3에서는 남학생들은 고등학생과 가정환경 간의 연관성이 있는 것으로 나타났고, 여학생들은 동아리, 대학원, 아르바이트 등 좀 더 다양한 활동들이 규칙에 포함되는 것을 확인하였다. 면접을 진행한 연구자는 여학생들에게서 나

타나는 ‘동생들’에 대한 잦은 언급이 많이 연구참여자들의 가족에 대한 책임과 부담감을 추정할 수 있게 하여준다고 해석하였다.

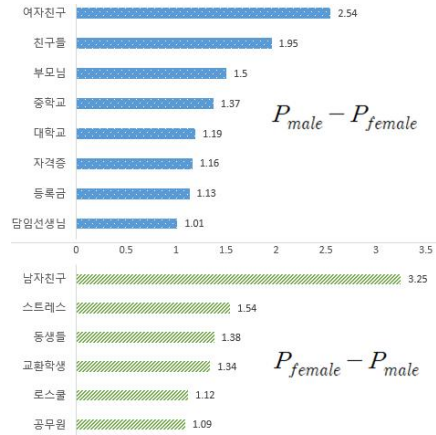


Fig. 2. Difference of words percentage by gender greater than one percent

Table 4. Example of association analysis results - male

|    | A             | ⇒ | B       | sup. | conf. | lift |
|----|---------------|---|---------|------|-------|------|
| 1  | {선생님, 중학교}    |   | {담임선생님} | 0.04 | 0.60  | 7.40 |
| 2  | {고등학교, 선생님}   |   | {담임선생님} | 0.05 | 0.50  | 6.17 |
| 3  | {등록금}         |   | {장학금}   | 0.03 | 0.47  | 4.38 |
| 4  | {장학금}         |   | {등록금}   | 0.03 | 0.28  | 4.38 |
| 5  | {담임선생님, 중학교}  |   | {선생님}   | 0.04 | 1.00  | 4.23 |
| 6  | {담임선생님}       |   | {선생님}   | 0.08 | 0.96  | 4.05 |
| 7  | {선생님}         |   | {담임선생님} | 0.08 | 0.33  | 4.05 |
| 8  | {실험실}         |   | {교수님}   | 0.03 | 0.64  | 4.05 |
| 9  | {교수님}         |   | {실험실}   | 0.03 | 0.19  | 4.05 |
| 10 | {고등학교, 담임선생님} |   | {선생님}   | 0.05 | 0.94  | 3.98 |

sup: support, conf: confidence

Table 5. Example of association analysis results - female

|    | A           | ⇒ | B       | sup. | conf. | lift |
|----|-------------|---|---------|------|-------|------|
| 1  | {담임선생님}     |   | {선생님}   | 0.04 | 0.83  | 3.98 |
| 2  | {선생님}       |   | {담임선생님} | 0.04 | 0.21  | 3.98 |
| 3  | {초등학교}      |   | {중학교}   | 0.03 | 0.45  | 3.92 |
| 4  | {중학교}       |   | {초등학교}  | 0.03 | 0.27  | 3.92 |
| 5  | {고등학교, 선생님} |   | {중학교}   | 0.03 | 0.38  | 3.33 |
| 6  | {실험실}       |   | {교수님}   | 0.04 | 0.57  | 3.24 |
| 7  | {교수님}       |   | {실험실}   | 0.04 | 0.22  | 3.24 |
| 8  | {고등학교, 친구들} |   | {대학교}   | 0.04 | 0.46  | 3.14 |
| 9  | {선생님, 중학교}  |   | {고등학교}  | 0.03 | 0.76  | 2.63 |
| 10 | {대학교, 선생님}  |   | {고등학교}  | 0.03 | 0.76  | 2.63 |

sup: support, conf: confidence

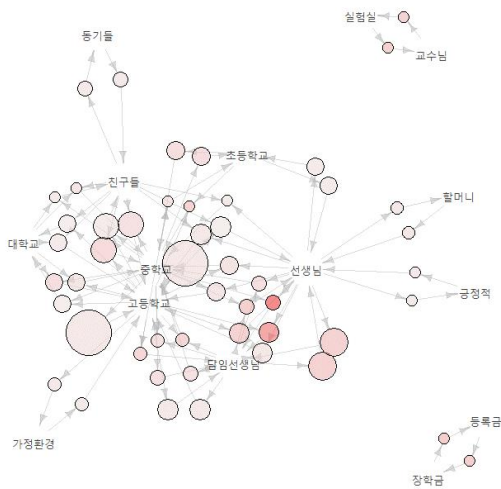


Fig. 3. Association analysis results - male

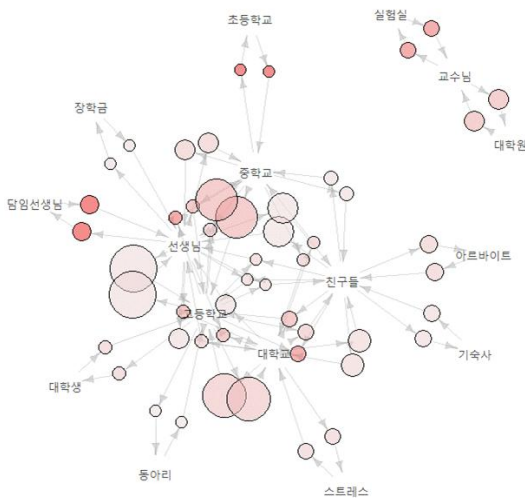


Fig. 4. Association analysis results - female

3.3 전공별 빈도 분석 및 연관 분석 결과

전공별 연구 대상자의 수를 고려하여 분석한 단어 빈도와 비율은 다음 Table 5와 같다. 인문사회경상계 전공 학생들의 경우 스트레스, 아르바이트, 남자친구, 재밌는, 교환학생에 대한 언급이 상대적으로 많은 것으로 나타났고, 이공계 전공 학생들은 고등학교, 교수님, 대학교, 중학교, 대학원, 부모님, 실험실에 대한 언급이 많은 것으로 나타났다. 교육계 전공 학생들은 다른 전공에 비해 선생님, 장학금, 동아리, 초등학교, 멘토링, 봉사활동 등의 단어들을 대화 중 많이 언급한 것으로 나타났다(Fig. 4).

Table 6. Percentage of words per major (%)

| words | H    | S    | E    | words | H    | S    | E    |
|-------|------|------|------|-------|------|------|------|
| 고등학교  | 6.70 | 7.58 | 5.75 | 담임선생님 | 1.30 | 1.51 | 1.82 |
| 선생님   | 4.42 | 5.16 | 9.26 | 교환학생  | 2.65 | 0.54 | 0.56 |
| 친구들   | 4.57 | 5.33 | 5.47 | 분위기   | 1.09 | 1.83 | 0.98 |
| 교수님   | 3.85 | 4.73 | 4.35 | 자격증   | 1.61 | 1.72 | 0.35 |
| 대학교   | 3.07 | 4.41 | 1.54 | 방학    | 0.99 | 1.45 | 1.33 |
| 중학교   | 2.60 | 3.34 | 2.53 | 동기들   | 1.25 | 0.86 | 1.40 |
| 대학원   | 0.21 | 4.41 | 3.58 | 멘토링   | 1.51 | 0.38 | 1.61 |
| 스트레스  | 3.07 | 2.21 | 2.39 | 대학생   | 0.78 | 1.24 | 1.05 |
| 기숙사   | 2.49 | 2.21 | 2.60 | 동생들   | 0.52 | 2.15 | 0.21 |
| 아르바이트 | 3.17 | 2.42 | 1.26 | 프로그램  | 1.30 | 1.08 | 0.28 |
| 남자친구  | 4.26 | 0.22 | 1.96 | 학생들   | 0.73 | 1.13 | 0.91 |
| 장학금   | 2.23 | 1.40 | 2.81 | 로스쿨   | 2.23 | 0.16 | 0.07 |
| 부모님   | 1.30 | 3.07 | 0.84 | 여자친구  | 0.94 | 0.48 | 1.40 |
| 동아리   | 1.61 | 1.51 | 2.39 | 자신감   | 1.09 | 0.48 | 1.12 |
| 선배들   | 1.56 | 1.99 | 1.75 | 자기소개  | 1.09 | 0.97 | 0.42 |
| 재밌는   | 2.34 | 1.02 | 1.82 | 공무원   | 2.18 | 0.00 | 0.07 |
| 실험실   | 0.00 | 4.79 | 0.00 | 봉사활동  | 0.78 | 0.32 | 1.47 |
| 초등학교  | 1.40 | 1.56 | 2.18 | 대외활동  | 1.66 | 0.43 | 0.07 |
| 할머니   | 0.94 | 0.65 | 3.58 | 등록금   | 1.25 | 0.70 | 0.21 |

H: Humanities and social sciences, S: Science and engineering, E: Education

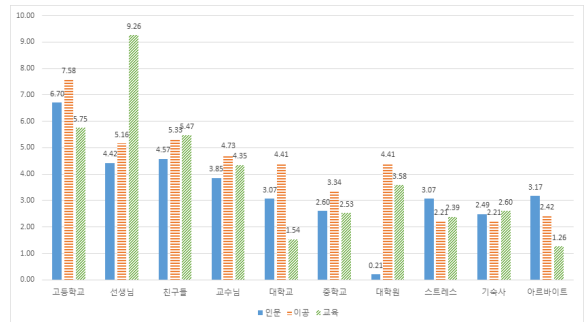


Fig. 5. Comparison of word frequency ratio by major

전공에 따른 단어 간의 관계를 살펴보기 위해 연관분석을 실행한 결과는 다음 Table 6과 같다. 분석 결과 향상도가 높은 상위 규칙은 사례 수가 적어 개개인의 특성의 영향을 받은 것처럼 보인다. 상위 50개 규칙을 기준으로 단어 간 관계를 살펴본 Fig. 5에서는 이공계 전공의 경우 대학원, 실험실, 교수님, 스트레스, 분위기 등의 단어 관련성이 다른 군집으로 나타났다. 교육계의 경우 개인별 특성이 반영되어 있지만 부모님, 할머니, 할아버지 등 가족과 관련된 단어들이 다른 단어와 관련 있는 것으로 나타났다.

Table 7. Example of association analysis results – humanities and social sciences

|    | A           | ⇒ | B       | sup. | conf. | lift |
|----|-------------|---|---------|------|-------|------|
| 1  | {담임선생님}     |   | {선생님}   | 0.04 | 0.82  | 5.93 |
| 2  | {선생님}       |   | {담임선생님} | 0.04 | 0.30  | 5.93 |
| 3  | {공무원}       |   | {로스쿨}   | 0.03 | 0.37  | 5.33 |
| 4  | {로스쿨}       |   | {공무원}   | 0.03 | 0.43  | 5.33 |
| 5  | {초등학교}      |   | {중학교}   | 0.03 | 0.55  | 4.44 |
| 6  | {중학교}       |   | {초등학교}  | 0.03 | 0.27  | 4.44 |
| 7  | {고등학교, 선생님} |   | {대학교}   | 0.04 | 0.60  | 4.05 |
| 8  | {고등학교, 친구들} |   | {중학교}   | 0.04 | 0.43  | 3.46 |
| 9  | {중학교, 친구들}  |   | {고등학교}  | 0.04 | 0.92  | 3.43 |
| 10 | {고등학교, 친구들} |   | {대학교}   | 0.04 | 0.46  | 3.14 |

sup.: support, conf.: confidence

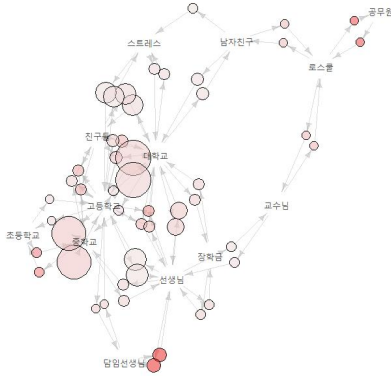


Fig. 6. Association analysis results – humanities and social sciences

Table 8. Example of association analysis results – science and engineering

|    | A             | ⇒ | B       | sup. | conf. | lift |
|----|---------------|---|---------|------|-------|------|
| 1  | {고등학교, 선생님}   |   | {담임선생님} | 0.04 | 0.41  | 7.17 |
| 2  | {담임선생님}       |   | {선생님}   | 0.05 | 0.94  | 5.46 |
| 3  | {선생님}         |   | {담임선생님} | 0.05 | 0.31  | 5.46 |
| 4  | {고등학교, 담임선생님} |   | {선생님}   | 0.04 | 0.92  | 5.33 |
| 5  | {교수님, 대학원}    |   | {실�험실}  | 0.04 | 0.67  | 4.08 |
| 6  | {대학원, 실험실}    |   | {교수님}   | 0.04 | 0.67  | 3.85 |
| 7  | {초등학교}        |   | {중학교}   | 0.04 | 0.52  | 3.63 |
| 8  | {중학교}         |   | {초등학교}  | 0.04 | 0.24  | 3.63 |
| 9  | {실험실}         |   | {교수님}   | 0.10 | 0.59  | 3.40 |
| 10 | {교수님}         |   | {실험실}   | 0.10 | 0.56  | 3.40 |

sup.: support, conf.: confidence

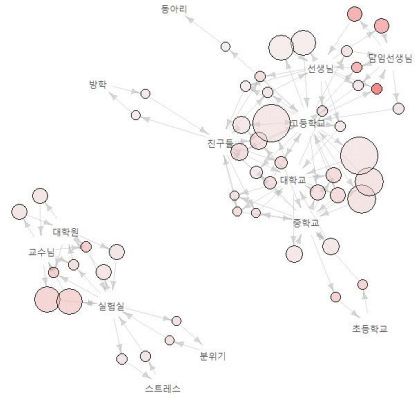


Fig. 7. Association analysis results – science and engineering

Table 9. Example of association analysis results – education

|    | A             | ⇒ | B       | sup. | conf. | lift  |
|----|---------------|---|---------|------|-------|-------|
| 1  | {교수님, 대학원}    |   | {사회학}   | 0.03 | 0.46  | 12.92 |
| 2  | {사회학}         |   | {대학원}   | 0.04 | 1.00  | 7.00  |
| 3  | {대학원}         |   | {사회학}   | 0.04 | 0.25  | 7.00  |
| 4  | {교수님, 사회학}    |   | {대학원}   | 0.03 | 1.00  | 7.00  |
| 5  | {고등학교, 담임선생님} |   | {중학교}   | 0.04 | 0.80  | 5.60  |
| 6  | {고등학교, 초등학교}  |   | {중학교}   | 0.04 | 0.73  | 5.09  |
| 7  | {할아버지}        |   | {할머니}   | 0.03 | 0.75  | 5.07  |
| 8  | {할머니}         |   | {할아버지}  | 0.03 | 0.21  | 5.07  |
| 9  | {선생님, 중학교}    |   | {담임선생님} | 0.05 | 0.43  | 4.67  |
| 10 | {고등학교, 중학교}   |   | {담임선생님} | 0.04 | 0.38  | 4.15  |

sup.: support, conf.: confidence

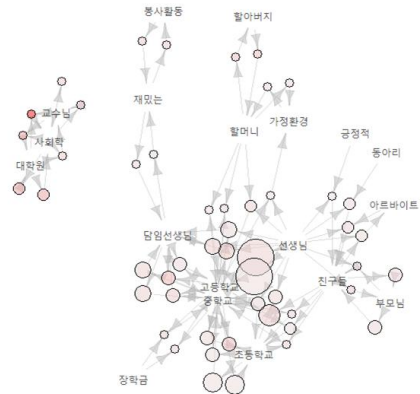


Fig. 8. Association analysis results – education

#### 4. 결론 및 논의

이 연구는 저소득층 대학생들의 경험이 성별과 전공

계열에 따라 어떠한 내적 차이가 있는지를 텍스트마이닝 기법을 활용하여 탐색적으로 분석해보고자 하였다. 이 연구를 통해 나타난 결론은 다음과 같다.

첫째, 연구 대상의 대학생들은 인터뷰 과정에서 이전의 학업 경험을 회상할 때, 가정환경, 부모님 등의 단어를 언급하는 경우가 많았으며, 대학교 생활과 관련하여 긍정적, 현실적, 어려움, 심리적, 안정적 등의 감정적인 단어를 언급하는 경우가 많은 것으로 나타났다. 이를 통해 대학생들이 겪는 대학생활의 적응 과정에서 연구 대상들은 가정환경으로 인한 어려움이나 미래에 대한 소망이 내재되어 있는 것으로 보인다. 둘째, 인터뷰 시기에 따라 초반에는 초·중·고등학교와 친구들에 대한 언급이 많았던 것에 비해, 중반에는 교수님, 선배들, 실험실, 아르바이트, 공무원, 봉사활동 등 대학 생활 또는 진로와 관련된 단어들의 비율이 높아지는 것으로 나타났다. 이는 2회의 인터뷰 과정에서 대학생활에 점차 적응해 가며, 과거의 학업 경험에서 대학 생활과 진로에 관한 이야기가 인터뷰 내용의 많은 부분을 차지하게 된 것으로 보인다. 셋째, 성별 분석 결과 남학생들은 여학생들에 비해 과거 경험에 대한 언급이 더 많았고, 상대적으로 여학생들은 교환학생, 공무원 등 진로와 관련된 단어의 언급이 많은 것으로 나타났는데, 이는 남학생들이 군대를 다녀오는 동안 여학생들은 진로와 관련된 활동을 시작하였고, 또한 이 연구대상에서는 남학생들이 여학생들에 비해 기존의 대인 관계를 유지하는 경향이 있기 때문인 것으로 보인다. 넷째, 전공별 분석 결과 인문사회경상계 전공 학생들의 경우 스트레스, 아르바이트, 교환학생에 대한 언급이 상대적으로 많은데 비해, 이공계 학생들의 경우 실험실에 대한 언급이 등장하였으며, 실험실, 교수님, 스트레스, 분위기와 같은 단어들이 군집을 이루는 것으로 나타났다. 교육계 전공 학생들의 경우 선생님, 멘토링, 봉사활동 등의 단어들을 더 많이 언급하는 경향이 있는 것으로 나타났는데, 이는 각 전공별 특성을 어느 정도 반영한 것으로 보인다.

위와 같은 분석결과에 따르면, 동일한 저소득층 배경의 대학생이라도 전공계열과 성별에 따라 진로준비를 본격화하는 시기와 주요 관심과 생활영역, 관계의 특성에서 질적 차이가 나타난다. 이는 진로준비과정에서 이들의 경험세계를 반영하여 단순한 재정적 지원을 넘어서는 보다 세분화된 지원제도가 필요함을 보여준다.

이 연구에서는 텍스트마이닝 기법인 핵심단어의 빈도

분석과 연관분석을 통해 저소득층 대학교 장학생들의 대학생활과 진로 준비와 관련된 단어들의 빈도와 관계를 살펴보고자 하였다. 이는 비정형 데이터인 텍스트를 활용하여 대학생들의 진로준비과정을 단어 간 관계로 살펴봄으로써, 기존의 진로준비 연구에서 사용되는 사례연구 또는 합의적 질적 연구(Consensual Qualitative Research, CQR) 방식이 아닌 데이터마이닝 기법을 탐색적으로 적용해보고, 이를 신뢰도, 향상도와 같은 지수 및 시각화로 확인하였다는 점에서 이전의 연구들과 차별점이 있다. 또한 텍스트마이닝 기법을 활용한 시각화와 연관분석 방법이 후속 연구를 위한 기초자료로 활용될 수 있으며, 질적연구의 연구자 간 합의점이라는 주관성을 단어 간 관계로 보완할 수 있을 것이라는 점에서 의의가 있을 것으로 기대된다. 그러나 연구 대상이 장학생에서 표집되어 저소득층 대학생 일반을 대표하기에는 한계가 있으며, 단어 간 관계를 파악할 때 인터뷰 진행자의 개념 지도를 함께 확인할 수 없다는 점에서 제한점이 있다. 추후 연구에서는 이를 보완하여 인터뷰의 질문 내용을 통제하거나 인터뷰 진행자의 발화를 분석에 함께 고려하고, 더 다양한 사례를 수집하고, 일반 대학생 집단과 비교해봄으로써 보다 깊이 있는 단어 간 관계를 파악하여 진로준비과정에 대해 탐색해 볼 것을 제안한다.

## REFERENCES

- [1] H. R. Min. (2003). A Study on the Experiences and Costs of the Private Tutoring of University. Students for Preparing Employment. *Korean Journal of Sociology of Education*, 13(3), 133-149.
- [2] J. S. Jung & H. H. Kim. (2009). Analysis on the Influential Factors on Private Tutoring Expenditure of Korean College Students. *The Journal of Economics and Finance of Education*, 18(3), 89-122.
- [3] H. B. Park & S. S. Kim. (2011). An Analysis on The Effects of Individual Background, Employment Efforts, and University Selectivity on Labor Market Outcomes of College and University Graduates. *Korean Journal of Sociology of Education*, 21(3), 77-98.
- [4] H. J. Chu. (2012). Individual and College Effects on New Graduates. *The Korean Journal of Educational Administration*, 30(1), 603-626.
- [5] B. J. Kim & H. J. Seo. (2013). Analysis on Student and University Variables Affecting First Employment of



College Graduates. *The Journal of Economics and Finance of Education*, 22(4), 243-268.

[6] K. H. Park, J. S. Kim, C. H. Kim, J. Y. Namkung, S. J. Beak, H. J. Yang, S. S. Kim, W. J. Kim, B. W. Ha & K. Y. Han. (2017). *Comprehensive Analysis of Current Situation of Educational Gap in Korea*, Korea Educational Development Institute. Chungcheongbuk-do: KEDI(RR2017-17).

[7] J. K. Ahn & H. J. Bae. (2011). The effects of working while in college on the labor market transitions and outcomes. *Korean Journal of Sociology of Education*, 21(4), 149-180.

[8] J. S. Yang. (2014). *Graduation Status and Labor Market Performance of 4-year college graduates*. KRIVET Issue Brief, 56. Seoul: KRIVET.

[9] J. H. Lee & J. E. Chae. (2013). Low-income Students' Adaptation to University. *Journal of Lifelong Learning Society*, 9(1), 1-27.

[10] H. N. Jung. (2016). Differences in worker characteristics and labor market performance by types of work experience at school. *Employment issues*, 9(5), 78-87.

[11] J. H. Lee & M. H. Hwang. (2017). Case Study on Career Preparation Experiences of University Students from Low-Income Families. *Journal of Lifelong Learning Society*, 13(3), 57-83.

[12] J. S. Shim. (2011). Analysis of Conflict Frames Using Semantic Network Analysis. *Korea Public Administration Journal*, 20(2), 183-212.

[13] D. H. Goh & R. P. Ang. (2007). An introduction to association rule mining: An application in counseling and help-seeking behavior of adolescents. *Behavior Research Methods*, 39(2), 259-266.

[14] G. Shmueli, N. R. Patel & P. C. ruce. (2011). *Data mining for business intelligence: Concepts, techniques, and applications in Microsoft Office Excel with XLMiner*. New Jersey: John Wiley and Sons.

[15] R. Srikant & R. Agrawal. (1996, June). Mining quantitative association rules in large relational tables. *Acm Sigmod Record*, 25(2), 1-12.

이 지 혜(Lee, Jihye)

[정회원]



- 2000년 8월 : 서울대학교대학원 교육학과 (교육학박사)
- 2005년 3월 ~ 현재 : 한림대학교 교양기초교육대학 교수
- 관심분야 : 평생교육, 교육복지
- E-Mail : jihye@hallym.ac.kr

이 신 혜(Lee, Shinhye)

[정회원]



- 2014년 2월 : 카이스트 수리과학과(이학사)
- 2016년 8월 : 서울대학교 교육학과(교육학석사)
- 2017년 9월 ~ 현재 : 서울대학교 교육학과 박사과정
- 관심분야 : 교육통계 · 측정 · 평가
- E-Mail : lee.sinhye@snu.ac.kr