Korean Journal of General Education 2021. 10. Vol. 15, No. 5, PP. 169-183 https://doi.org/10.46392/kjge.2021.15.5.169 pISSN 1976-3212 eISSN 2714-1101 www.kagedu.or.kr

교양 교육 연구

Korean Journal of General Education

LDA 토픽 모델링을 활용한 대학 교양영어 연구동향 분석

박은희

부경대학교 강사, jasminepark7@gmail.com

초록

본 연구는 대학 교양영어 교육의 연구 현황을 분석하고 향후 교양영어 수업과 연구 방향에 대한 통찰을 얻고자 실시되었다. 분석기간은 2001년에서 2020년, 20년의 기간이었으며, 분석대상 논문은 313편이었다. 분석 방법은 연구 문제에 따라 빈도분석, LDA(잠재 디리클레 할당) 토픽모델링, 시계열 분석을 활용하였다. 분석 결과는 다음과 같다. 첫째, 빈도 분석 결과 대학 교양영어 관련 논문은 지난 20년간 양적으로 크게 증가하였으며, 연구 주제도 점차 다양해졌음이 확인되었다. 둘째, LDA 토픽모델링 분석 결과 교양영어 교육의 주제는 10가지로 분류되었다. 주제명은 '교육과정 및 수준별 수업주제1)', '학습자 정의적 요인(주제2)', '평가 및 학습 전략(주제3)', '교수자 요인(주제4)', '어휘·문법학습 및 쓰기활동(주제5)', '특수목적영어 프로그램(주제6)', '교수학습 방법(주제7)', '웹 기반 수업(주제8)', '학습자 중심 교육(주제9)', '교재 분석 및 기타(주제10)'였다. 셋째, 시계열 분석 결과 '학습자 정의적 요인(주제2)', '특수목적영어 프로그램(주제6)', '학습자 중심 교육(주제9)', 세 가지가 상승 주제인 것으로 드러났다. 이 세 주제는 모두 학습자의 동기, 학습자의 요구, 학습자가 중심이 되는 수업 활동이라는 점에서 모두 학습자 요인과 관련 있는 주제임을 알 수 있었다. 이 분석을 근거로 향후 학습자 요인과 관련된 연구가 활발히 진행될 것으로 전망되었다. 본 연구의 의의는 빅데이터 기반의 LDA 토픽 모델링 기법으로 교양영어 교육 연구동향을 분석하고 향후 연구 동향을 전망함으로써 관련 연구자들에게 객관적이고 유용한 정보를 제공하고 있다는 점이라고 생각된다.

주제어: 교양 영어, 대학영어, 연구 동향, 주제 분석, LDA 토픽 모델링, 잠재 디리클레 할당

^{1.} 서론

^{2.} 이론적 배경

^{3.} 연구 방법

^{4.} 연구 결과 및 논의

^{5.} 결론

1. 서론

영어는 이제 의사소통을 위한 국제어(international language)로서의 위상을 넘어 전 세계인들이 사용하는 세계어(world language)로서 자리 잡아 가고 있다. 세계어로서의 영어는 나라와 나라, 사람과 사람을 연결하여 세계를 하나의 공동체로 만드는데 크게 기여하고 있다. 이러한 현상은 전 세계인들과 동시적으로 정보를 공유하는 SNS의 시대를 맞아 더욱 가속화되는 것처럼 보인다. 영어사용능력이라는 시대적 요청에 맞추어 국내 대부분의 대학들은 필수 교양영어 과목을 개설하거나 다양한 교양영어 강좌를 선택 과목으로 개설함으로써 재학생들이 영어 의사소통 역량을 갖춘 글로벌 인재로 성장할 수 있도록 노력하고 있다.

우리나라 대학들이 교양영어 수업을 실시한 1960년대이래 교양영어 교육과정은 제2언어 습득과 관련한 담론의변천과 더불어 변화를 거듭해왔다. 교양영어 교육의 초창기에는 영미 문학작품의 원문을 읽으며 '교양'을 강조하는 강독 위주의 교육이 실시되었다. 이러한 강독 위주의교양영어 교육은 1970년대 후반 의사소통 교수법이 도입되고 실용영어에 대한 필요성이 제기되면서 변화의 국면을 맞이하였다. 1990년대에 이르러서는 영어 의사소통능력을 강조하는 말하기 중심의 실용영어 교육이 강화되고,일반 영어, 실무 영어, 전공 영어 등과 같은 교양영어 과목의 확대가 이루어졌다. 2000년대 이후에는 이러한 경향이더욱 강조되어 듣기와 말하기를 포함한 영어의 4가지 언어 영역을 모두 통합적으로 증진시키는 방향으로 교양영어 교육과정의 변화가 이루어졌다(김성혜, 임자연, 2013;송미정, 박용예, 2004).

그간 교양영어 수업과 관련한 많은 연구들이 발표되어 교육현장에 기여하였으며, 교양영어의 효과적인 운영과 방향 모색을 위한 거시적 관점의 연구들도 다수 발표되었다. 교양영어의 발전과 관련한 연구들이 공통적으로 제안하고 있는 방안을 보면, 교양영어 전담 기구 및 전담 교원확보, 교양영어 교육의 구체적인 목표 설정, 교양영어 교육의 성과를 평가 할 수 있는 시스템의 도입, 선택 교양영어 과목의 확대 등 이었다(김상수, 2009; 김성혜와 임자연, 2013; 방영주, 2004; 조정순, 2002). 지금까지 교양영어수업을 통한 많은 실험 연구들과 교양영어 교육을 연구자들의 제안으로 대학 교양영어의 발전이 이루어져왔다는

것은 자명하다할 것이다.

하지만 지금까지 교양영어과 관련한 많은 연구들이 특 정을 대학을 대상으로 한 사례 연구인 점, 많은 연구들이 설문조사 결과를 분석 데이터로 삼고 있는 점, 연구 동향 분석의 경우 연구자 중심의 분류 방식인 점 등을 고려해볼 때 새로운 접근 방법의 필요성이 제기된다. 현재 우리는 새로운 언어, 빅데이터의 시대를 살고 있기 때문이다. 조 성준(2019)은 '빅데이터가 인사이트로 바뀌는 과정은 재 료가 요리로 바뀌는 과정'(p.14)과 같다고 하였다. 우리는 매일 데이터를 생산하고 소비하고 있으며, 여러 분야에서 그 데이터들은 분석되고 활용된다. 마찬가지로 학술 연구 분야에서도 논문과 같은 비정형 텍스트 데이터를 분석할 수 있는 프로그램이 등장하여 적극적으로 활용되고 있다. 대학 교양영어의 효과적 운영이 절실한 작금의 현실에 비추어 볼 때, 빅데이터 기반 분석 방법을 활용한 교양영 어 교육에 대한 통찰이 필요하다고 생각된다. 이에 본 연구에서는 비정형 텍스트 데이터 분석 방법인 토픽 모델 링(Topic modeling) 기법을 활용하여 지금까지 이루어진 교양영어 연구의 동향을 읽고, 향후 연구방향을 제시함으 로써 교양영어 연구와 교육현장에 기여하고자 한다.

2. 이론적 배경

2.1 교양영어 교육 현황 및 발전 방안 연구

본 연구의 목적에 의거하여 지금까지 교양영어 교육 현황 및 발전 방안을 분석한 연구를 간단히 살펴보면 다음 과 같다. 조정순(2002)은 전국 대학의 교수자들(N=111) 와 학생들(N=825명)을 대상으로 교양영어 수업에 대한 설문조사를 실시하고 그 결과를 보고하였다. 교양영어 프로그램의 효율성 향상을 위한 조건으로 교수자들이 중 요하게 생각하는 요소는 '의무이수 학점 확대와 수강 연 한 연장'이 있었으며, 학생들이 중요하게 생각하는 요소 로는 '학교가 주관하는 영어권 국가로의 어학연수 기회 확대'가 있었다. 교수자들과 학습자들이 공통적으로 중요 하게 생각하는 요소는 '다양한 강좌개설을 통한 학생의 수강선택 폭 확대'와 '우수한 강사진의 확보'인 것으로 나타났다. 이 연구는 많은 수의 설문 데이터를 수집한 점, 교수집단과 학생집단의 의견을 모두 분석대상으로 삼고 있다는 점에서 수많은 후속연구에서 인용되었다. 교양영어 프로그램의 효과적 운영을 위해 학습자 집단과 교수자 집단 모두 다양한 교양영어 프로그램의 개설과 우수한 강사진 확보의 필요성에 대해 공감하고 있음을 분석결과로 제시한 점은 이 연구의 중요한 의의 중 하나로 생각된다.

방영주(2014)는 필수 교양영어를 이수한 학생들과 교양영어 선택과목을 수강하는 학생들(N=345)을 대상으로 설문조사를 실시하고 분석 결과를 보고하였다. 학생들은 강좌 당 학생 수를 묻는 질문에 대해, 회화 과목의 경우 20명 이내, 영어독해의 경우 25-30명 내외가 적당하다고 응답하였다. 학생들은 교양영어 프로그램의 효율적 운영과 개인 편차로 인한 교육 문제 해소를 해결할 수 있다는 점에서 수준별 수업에 대해 긍정적 반응을 보였으며 영어졸업인증제의 필요성에 대해서도 어느 정도 수긍하는 것으로 나타났다. 이와 함께 학생들은 좀 더 다양한 교양영어 선택과목을 희망하고 있다는 점도 보고되었다.

김성혜, 임자연(2013)은 교양영어 프로그램 개선 방안을 마련하기 위해 인터뷰와 설문조사를 실시하고 그 결과를 보고하였다. 인터뷰 대상자는 서울 소재 대학 교양영어 운영 담당자(N=8)였으며, 설문 대상자는 전국 대학의 교양영어 담당자(N=36)들이었다. 연구자들은 분석을 통해 교양영어 교육의 문제점을 세 가지로 지적하였다. 첫째, 교양영어 교육이 '실용영어 능력의 함양'이라는 다소 광범위한 목표를 제시하고 있는 점, 둘째, 설정된 목표에 대한 평가가 제대로 이루어지지 않고 있다는 점, 셋째로는 교양영어 교수자의 채용 및 관리가 체계적이지 않다는 점을들었다. 연구자들은 이런 문제를 해결하기 위해서 교양영어를 담당하는 전담인력을 확보하는 것이 현실적인 측면에서 가장 중요한 개선 방안이 될 것으로 전망하였다.

지금까지 살펴본 연구 결과들을 요약하면, 교양영어 교육의 발전방안으로 '교양영어 교육의 구체적인 목표 설정'과 '목표에 맞는 평가도구 도입', '전담 인력의 필요성' 및 '우수한 강사진 확보', '다양한 강좌 개설' 등을 제시하고 있다는 점, 세 연구 모두 연구 방법에 있어서 설문조사 결과를 토대로 분석하였다는 점을 들 수 있겠다.

2.2 교양영어 연구동향 분석 연구

대학 교양영어의 연구동향을 분석한 연구는 많지 않은

편이며, 주요 연구로 이연숙(2014)과 하명애(2017)가 있 다. 이연숙(2014)은 교양영어 관련 연구들을 분석하고 대 학영어 프로그램의 주제를 여섯 가지로 분류하였다. 주제 1은 '요구분석/의식조사', 주제2는 '대학영어교육의 목표 와 방향 (① 일반영어, ② ESP)', 주제3은 '평가인증시험', 주제4는 '교재 개발', 주제5는 '교육방안 개발 (① 수준별 수업, ② 사이버/온라인/멀티미디어 활용 학습, ③ 실용영 어 교수법, ④ 학습동기 및 전략, ⑤ 수업 평가)', 주제6은 '전체 과정평가'였다. 연구자는 '요구분석/의식조사(주제 1)' 연구의 빈도가 가장 높다고 보고하였으며 이는 교육의 중심이 교육의 공급자로부터 수요자로 변환되고 있는 추 세를 반영하는 것이라고 설명하였다. 한편 '평가인증시험 (주제3)', '교재개발(주제4)', '전체 과정평가(주제6)'에 대한 연구가 미흡하며 특히 대학영어를 통합적인 프로그 램으로 인식하고 분석한 연구가 많지 않다고 지적하였다. 이 연구에서 연구자는 주제별로 대표 논문을 몇 가지씩 소개하고 있으며 중복으로 분류된 논문 포함 총 79편의 논문을 소개하였다.

하명애(2017)는 1968년부터 2017년 상반기까지 교양 영어 관련 연구 227편을 대상으로 연구 동향을 분석하였 다. 연구자는 교양영어 연구 주제를 다섯 가지 기준으로 분류하였다. 주제1은 '교육과정 및 교육방안 (① 교육과정 방안, ② 수준별 수업)', 주제2는 '교수학습방법 (① 교수 학습법, ② 교재·자료, ③ 멀티미디어 활용, ④ 수업목적 과 내용)', 주제3은 '학습자 요인 (① 요구분석, ② 인식 조사, ③ 학습심리)', 주제4는 '교수자 요인', 주제5는 '평 가'였다. 연구자는 전체 기간 동안 가장 많이 연구된 주제 는 '교육과정 방안(주제1)', '교수학습법(주제2)', '평가 (주제5)'이며, 1968~1989년까지는 '교육과정 및 교육방 안(주제1)'에 대한 연구가 가장 많았던데 비해 2000년대 이후에는 '수준별 수업(주제1의 소주제)'과 '멀티미디어 활용(주제2의 소주제)'연구가 가장 많았다고 분석하였 다. 연구자는 교양영어 연구의 특징에 대해 첫째, 2000년 대 이후 관련 연구가 급증하고 있다는 점, 둘째, 연구 주제 가 점점 세분화되고 다양해지고 있다는 점을 들었다. 이 연구는 교양영어 연구가 시작된 시점부터 연구시점인 2017년까지의 연구를 대상으로 교양영어 연구들을 모두 수집하여 주제별로 분류하고 연구동향을 분석하고 있다 는 점에서 의의가 있다고 하겠다.

교양영어 연구 동향을 분석한 두 연구의 분석 결과를

비교해보면, 이연숙(2014)의 경우 '요구분석/의식조사' 관련연구가 가장 많이 이루어졌고 '평가인증시험', '교재 개발', '전체과정 평가'에 대한 연구가 부족하다고 분석한 반면, 하명애(2017)는 '교육과정 방안'에 대한 연구가 가 장 많고 그 다음으로 '교수학습법', '평가'라고 분석하였 다. 즉, 논문 출판 시점이 2014년과 2017년으로 3년 밖에 차이가 나지 않음에도 불구하고 두 연구의 분석결과에 큰 차이가 있는 것이다. 이러한 결과의 차이는 두 가지의 원인에 의한 것으로 생각해볼 수 있는데 첫째, 두 연구의 주제 분류가 같지 않다는 점이다. 예를 들면 평가의 경우 이연숙(2014)은 주제로 '평가 인증시험', '전체 과정 평가' 가 있고 주제 '교육방안 개발' 안에 소주제로 '수업평가' 가 있다. 이처럼 이연숙(2014)의 분류는 평가 관련 주제가 세 가지로 세분화 되어있는 반면, 하명애(2017)의 경우 '평가' 관련 항목은 하나로 통합되어 있음을 알 수 있다. 기준이 다르면 분석결과가 달라지는 것은 당연한 결과라 하겠다. 두 번째, 두 연구 모두 연구 동향을 분석함에 있어 주제 분류 기준을 먼저 설정하고 난 후 각 논문이 어느 주제에 해당하는지 분류하였기 때문에 발생하는 차이일 수 있다는 것이다. 실제로 하나의 연구는 하나 이상의 주제를 포함하고 있을 가능성이 많음에도 불구하고 하나 의 논문을 하나 혹은 두 개의 주제로 분류하는 방식을 택함으로써 전체 데이터가 가지는 모든 주제 속성이 다 추출되지 않았을 가능성을 생각해 볼 수 있다는 것이다.

지금까지 살펴본 내용을 바탕으로 본 연구의 필요성을 피력하고자 한다. 첫째, 여러 연구(김성혜와 임자연, 2013; 방영주, 2014; 이연숙, 2014)에서 지적한 바와 같이 대학 교양영어 관련 연구 대부분이 수준별 학습, 온라인 학습 등 한정적인 연구문제를 다루고 있거나 특정 대학의 사례 를 대상으로 하고 있는 경우가 많고 교양영어 교육에 대한 거시적 관점의 연구동향 분석과 전망에 대한 연구가 상대 적으로 부족하다는 점이다. 둘째, 교양영어 교육의 연구 동향을 분석한 연구들의 경우, 연구자가 먼저 주제를 설정 한 후 논문들을 분류함으로써 연구 간 주제 분류와 분석 결과에 차이가 발생하였다는 점(이연숙, 2014; 하명애, 2017)이다. 따라서 교양영어 관련 연구들을 광범위하게 수집하고, 연구들에 잠재되어 있는 모든 주제를 추출하여 분석하고, 주제별 연구 전망을 제시함으로써 좀 더 객관적 이고 유용한 연구동향 분석 결과를 도출해야할 필요성이 제기된다. 이러한 필요에 근거하여 본 연구의 연구 질문을

다음과 같이 설정하였다.

첫째, 2001년부터 2020년까지 교양영어 연구의 학술지 게재 현황은 어떠한가?

둘째, 동 기간 교양영어 교육의 연구 주제는 어떠한 것들이 있는가?

셋째, 교양영어 교육 연구 주제들의 향후 전망은 어떠 한가?

3. 연구 방법

3.1 분석 대상

데이터 수집을 위해 넷마이너(Netminer) 4.4.3 확장 프로그램인 비블리오 데이터 콜렉터(Biblio Data Collector)를 활용하였다. 비블리오 데이터 콜렉터로 특정 분야의국내 논문을 수집하는 경우 KCI에 등록된 논문들이 모두수집되며 상세수집 기능을 활용하면 분석 대상이 되는연구 논문들의 논문 제목, 저자명, 저널명, 출판학회명, 출판일자, 키워드, 초록과 같은 서지정보를 수집할 수 있다.

데이터 수집을 위해 검색 기간을 1900년에서 2020년으로 설정하여 KCI에 등록된 교양영어 관련 논문을 수집한결과, 수집된 전체 논문 중 1900년대 논문은 2편(이효웅, 1996; 김종보, 1997) 뿐이었으며, 나머지는 모두 2000년대 이후 출판된 논문이었다. 두 편으로 1900년대 교양영어 교육의 연구동향을 파악하기는 무리가 있다고 판단되어, 연구기간을 2001년에서 2020년, 20년의 기간으로 설정하기로 하였다.

데이터 수집을 위해 'college English', 'university English', 'general English', 'freshman English', 'freshmen English', '교양영어', '실용영어', '대학영어', '신입생 영어' 등과 같은 여러 가지 검색어로 논문을 수집해보았으며, 수집결과 대학에서 이루어지는 대부분의 영어 관련 연구가모두 수집되어 교양영어 교육에 한정한 논문 수집에 한계가 있음을 알 수 있었다. 이에 본 연구에서는 '교양영어', '교양 영어', '신입생 영어'로 한정하여 데이터를 수집하였다. '신입생 영어'라는 키워드를 최종 키워드로 포함시킨 이유는 대학 신입생을 대상으로 하는 필수 교양영어과목에 대한 연구를 반드시 포함하고자 했기 때문이다. 이 방법으로 수집된 연구들 중에서도 연구 목적과 맞지

않는 연구들, 예를 들면 영어 전공 학생들을 대상으로 한 연구, 대학원생을 대상으로 한 연구(예: 최원경, 2020), 교양영어와 함께 다른 교양과목(예: 교양국어 등)을 포함 하여 분석한 연구 등은 연구 대상에서 모두 제외하였다. 여러 번의 수집과 필터링 과정을 통해 선정된 최종 분석대 상 논문은 313편이었다.

따라서 본 연구의 분석대상은 2001년부터 2020년까지 KCI에 등재된 논문들 중에서 논문 초록, 논문 제목, 저자 키워드 중 하나 이상에 '교양영어', '교양 영어' 혹은 '신입생 영어'가 포함된 313편의 논문이 된다. 최종 선정된 313편 논문들의 서지정보를 확인해본 결과, 이 중 13편의 논문들은 영문 초록이 누락되어 있었다. 따라서 영문초록이 포함되어 있지 않은 13편의 논문은 주제 분석 및 연구동향 분석에서 제외되었다. 따라서 연구문제 1의 분석대상은 313편, 연구문제 2와 연구문제 3의 분석 대상은총 300편의 논문이 된다.

3.2 분석 방법

본 연구의 데이터 분석을 위해 넷마이너 4.4.3 프로그램을 사용하였다. 데이터 분석을 위해 전처리를 실시하였으며 연구문제 1, 2001년~2020년 발행 논문 수 분석은 빈도 분석을 실시하였고, 연구문제 2, 주제 분석은 LDA(Latent Dirichlet Allocation, 잠재 디리클레 할당, 이하 LDA) 토픽 모델링 분석을 활용하였으며, 연구문제 3, 주제별 연구동향 분석은 시계열 분석(Time series analysis)을 통해이루어졌다.

3.2.1 데이터 전처리(Data Preprocessing)

넷마이너는 형태소 분석기능이 탑재되어 있어 분석대상 형태소를 지정할 수 있다. 넷마이너를 이용한 텍스트분석의 경우 명사로 지정하는 것이 일반적이며(박은희, 2021; 원용국, 김영우, 2021; 임수민, 윤회정, 방담이, 2020; 하근희, 2021), 본 연구에서도 명사를 분석 형태소로 지정하였다. 수집된 논문을 대상으로 LDA 분석을 실시하기위해 전처리(Preprocess) 과정을 거쳤다. 넷마이너에서 전처리란 정확한 데이터 분석을 위해 사전을 만들어 처리하는 과정을 말한다. 사전의 종류는 세 가지이며 유사어사전(Thesaurus)→지정어 사전(Defined words)→제외어

사전(Exception list)의 순서로 입력 및 처리된다.

첫 단계, 유사어 사전이란, 같은 의미로 사용되는 단어 들을 유사어로 묶어 같은 단어로 인식되도록 처리해주는 것을 말한다. 예를 들어 'ESP'와 'English for specific purposes'를 하나의 단어로 인식하도록 하는 방식이다. 두 번째 단계, 지정어 사전이란, 띄어쓰기가 포함된 두 단어를 하나의 단어로 인식하도록 처리해주는 것을 말한 다. 예를 들어 'Business English'가 'Business'와 'English', 각각이 아닌 하나의 분석 단위, 'Business English'로 인식 되도록 지정하는 것이다. 세 번째 단계, 제외어 사전이란, 분석에 포함될 필요가 없는 단어가 제외되도록 처리하는 것을 말한다. 예를 들어, '교양 영어'는 모든 논문에 들어 있고 분석의 주제가 되므로 당연히 제외하여야 한다. 이외 에도 'data', 'discussion'과 같이 대부분의 초록 데이터에 들어있지만 분석에 의미를 부여하기 어려운 단어들은 모 두 제외어 사전에 포함시켰다. 전처리는 분석 결과의 신 뢰도와 정확성에 큰 영향을 미치는 중요한 단계이므로 여러 번의 반복 작업을 통해 최대한 정교하게 하고자 노력하였다.

3.2.2 LDA 토픽 모델링 분석

토픽 모델링이란, 문서들의 집합에서 '토픽(topic, 이하주제)'를 추출해내는 방법을 말하며 구조화되지 않은 대량의 텍스트를 분석하여 숨겨져 있는 주제 구조를 발견하고 범주화하는 통계적 추론 알고리즘을 일컫는다. 토픽모델링 기법 중 하나로 LDA가 많이 활용되며 최근 주제분석을 위한 많은 연구 분야에서 활발히 활용되고 있다. 본 연구의 분석 프로그램인 넷마이너도 토픽 모델링 방법으로 LDA가 적용되어 있다.

LDA는 Blei, Ng, Jordan(2003)이 처음 소개한 방법으로 "주어진 텍스트 집합의 숨겨진 주제 구조를 탐색하는 컴퓨터를 사용한 내용 분석 방법"(Maier 외, 2018: 93)으로 정의된다. LDA의 기본 가정은 문서는 주제들로 구성된 집합이며 주제의 속성은 단어로 표현된다는 것이다. LDA는 코퍼스 수준, 문서 수준, 단어 수준의 세 가지층위로 구성되며 특정 문서에서 주제와 단어를 반복적으로 선택하는 과정을 통해 문서 집합 전체의 주제와 각문서들의 주제를 추출해낸다. [그림 1]은 LDA를 처음으로 소개한 Blei, Ng, Jordan(2003)중 1저자인 Blei가 2012

년 자신의 논문(Blei, 2012)에서 LDA 토픽모델링을 그림 으로 나타낸 것이다.

[그림 1]은 문서 집합에서 전체 주제를 추출하고, 각 문서의 주제가 어떤 방식으로 포함되어 있는지를 분석하 는 과정을 나타낸다. 그림 왼쪽에 세로로 정렬된 네 개의 도형은 전체 문서 집합을 대상으로 추출된 주제(Topics) 가 4가지임을 보여준다. 하나의 도형은 같은 주제로 묶인 키워드들을 제시하고 있다(예를 들어, 맨 위 노란색 도형 에는 gene, dna, genetic ... 라는 키워드가 적혀있다). 그림 오른쪽에 있는 막대 그래프를 보면, 그림 속 예시 문서 (Seeking Life's Bare(Genetic) Necessities)에서 추출된 주 제가 3가지임을 알 수 있다. 그래프에서 노란색으로 표시 된 두 번째 막대가 가장 높은 것으로 보아 이 문서는 노란 색 도형에 적혀 있는 키워드가 가장 많이 포함된 문서임을 알 수 있는 것이다. 여기서 기억해야 할 점은 이 문서의 주제가 하나가 아닌 세 가지라는 점, 세 가지 주제에 해당 하는 키워드들이 모두 추출되어 전체 주제 분석에 이용된 다는 점, 그리고 하나의 문서는 가장 많이 포함된 주제로 분류(Topic proportions and assignments)된다는 점이다.

Blei(2012)의 [그림 1]은 LDA 토픽모델링 방법이 기존 연구동향 분석에서 사용되었던 연구방법과 어떤 차이가 있는지를 잘 드러내 보여준다. 첫째, 기존의 방식이 연구자가 주제를 먼저 설정하고 각 문서를 주제별로 분류한데 반해, LDA 토픽 모델링은 키워드 수집을 통해 모든 문서들의 숨겨진 주제를 모두 '드러나게' 함으로써 문서 집합

전체의 주제가 '데이터에 의해' 결정된다는 점이다. 둘째, 기존의 연구 동향 방식에서는 하나의 문서(논문)를 하나혹은 두 개의 주제로 분류하였다면, LDA 토픽 모델링은 각 문서에 내포된 모든 주제를 다 추출함으로써 하나의 문서가 N개의 주제를 얼마만큼 내포하고 있는가를 다보여준다는 점이다. 이러한 차이점으로 인해 LDA 토픽모델링은 연구자 중심 주제 분류방식보다 훨씬 더 정확하고 객관적인 분석이 가능하다는 점을 알 수 있다.

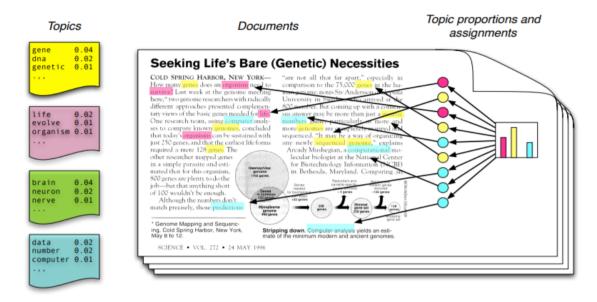
본 연구는 2001년에서 2020년 사이에 발표되어 KCI에 등재된 교양영어 교육 관련 연구 논문 313편 중 영문초록이 포함된 300편의 논문을 대상으로 LDA 분석을 실시하였으며, 수집된 키워드 및 실루엣 계수(Silhouette coefficient)라는 정량적 방법을 활용하여 주제 개수를 최종 결정하였다. 주제 개수 결정과 주제명 명명 과정에서 연구자의주관적인 판단을 배제하기 위해 영어교육 전공 학자 2명의 의견을 참고하였다.

4. 연구 결과 및 논의

4.1 학술지 게재현황

4.1.1 연도별 논문 수

분석대상인 논문 313편의 출판 연도를 기준으로 연도



[그림 1] LDA 토픽 모델링(Blei, 2012)

별 논문 수를 분석하였다. 2001년에서 2020년 사이 KCI에 등록된 교양영어 교육 관련 연도별 논문 수는 <표 1>과 같다. 2000년대에는 0편~11편 사이로 발표되던 논문 수가 2010년대 초반부터 크게 증가세를 보이는 것을알 수 있다. 특히, 2013년부터는 20편 이상으로 늘어나고, 2018년부터는 30편 이상으로 증가하고 있음을알 수 있다.

논문 발표 편수를 5년 단위로 분석해보면 시기별 차이 가 분명하게 드러난다. 표에 정리한 바와 같이, 2001년에 서 2005년까지의 총 논문 수는 26편, 연 평균 5.2편이었으 며 2006년에서 2010년까지 발표된 논문 수는 44편으로 연 평균 8.8편이었다. 2011년부터 2015년까지의 총 논문 수는 89편으로 연 평균 17.8편, 2016년부터 2020년 사이 의 논문 수는 154편, 연 평균 30.8편이었다. 이 분석결과를 보면 과거 2001년에서 2005년까지 5년 동안 발표되었던 논문 수(26편)보다 2018년 이후 한 해에 발표되고 있는 논문 수(31편~41편)가 더 많다는 것을 알 수 있다. 실로 교양영어 관련 연구 논문의 증가세가 엄청나다고 할 수 있겠다. 논문 수 증가의 원인으로는 첫째, 2000년대에 들어서면서 대학들이 사회적 변화와 학생들의 요구를 바 탕으로 교양영어 교육을 위해 노력한 점(성명희, 장형지, 김현지, 2007), 두 번째는 2000년을 전후로 영어교육 및 영어 관련 학술지, 교양교육 관련 학술지 등이 다수 창간 됨으로써 교양영어 교육 관련 출판 통로가 확대된 점 등이 주요한 이유가 될 것으로 보인다.

4.1.2 저자 분석

2001년에서 2020년 사이 교양영어 교육 관련 논문을 발표한 연구자를 대상으로 연구자별 논문 편수를 분석하 였으며 그 결과는 <표 2>와 같다. 표의 숫자는 공저인 경우 공저자 수와 상관없이 모든 저자를 각각 1편으로 집계한 수치이다.

< 포 2>에 제시된 바와 같이 교양영어 관련 논문을 발표한 저자 수는 총 292명이었다. 이 중에서 1편의 논문을 쓴 저자가 219명으로 가장 많았으며 전체 292명 중 75%를 차지하였다. 2편을 쓴 저자는 모두 42명으로 전체 저자수의 14.38%였다. 여기서 1편 혹은 2편을 쓴 저자 수를 모두 합하면 261명으로 전체의 89%에 해당된다. 3편을 쓴 저자는 19명, 4편을 쓴 저자는 6명이며 5편 이상을 쓴 저자는 6명이었다. 전체 논문 수가 313편인 상황에서 저자 수가 292명이라는 것은 교양영어 관련 논문을 저술한 연구자가 매우 넓게 분포되어 있다는 의미가 된다. 앞서 연도별 논문 수 분석에서 살펴본 바와 같이 2000년 대 이후 교양영어 관련 연구가 크게 증가한 것과 더불어 교양영어 연구자가 저변 확대된 것으로 볼 수 있다.

하지만 이를 달리 말하면, 교양영어를 본인의 관심 분야, 전문 연구영역으로 삼고 연구를 계속 이어가고 있는 연구자가 그만큼 적다는 의미로도 풀이 될 수 있을 것이다. 대부분의 대학에서 교양영어를 필수과목으로 지정하며 영어교육을 위한 노력을 경주하고 있음에도 불구하고

/п	1\	연도별	누므	亼
〈並	17	언노별	근근	7

년도	논문 수						
2001	0	2006	9	2011	7	2016	23
2002	5	2007	7	2012	16	2017	22
2003	2	2008	7	2013	22	2018	31
2004	8	2009	10	2014	24	2019	41
2005	11	2010	11	2015	20	2020	37
2001~2005	26	2006~2010	44	2011~2015	89	2016~2020	154

⟨표 2⟩ 저자별 논문 수 분석

논문 발표 수	저자 수	백분율
1편	219	75.00%
2편	42	14.38%
3편	19	6.50%
4편	6	2.05%
5편 이상	6	2.05%
합계	292	100.00%

/₩	3)	LDA	토픽부석	결과

주제	키워드 1	키워드 2	키워드 3	키워드 4	키워드 5	논문 수
1	course	questionnaire	level-differentiated	program	professors	56
2	motivation	difference	questionnaire	reading	anxiety	28
3	test	level-differentiated	strategy	score	TOEIC	27
4	questionnaire	NEST	interview	experimental group	control group	32
5	vocabulary	grammar instruction	test	error	type	27
6	program	curriculum	tutee	liberal arts	need	25
7	flipped learning	blended learning	classroom	interests	movie	21
8	course	content	professors	online class	online	20
9	activity	questionnaire	environment	classroom	role	35
10	textbooks	competence	knowledge	ability	development	29

교양영어 관련 전문 연구자의 수가 이토록 적다는 것은 매우 안타까운 일이 아닐 수 없다. 대학 교양영어 교육 개선방향을 발표한 여러 연구에서 교양영어 교육 전문 인력의 확보가 시급하다고 지적한 바 있으나(김성혜, 임자연, 2013; 조정순, 2002), 본 연구의 저자 분석 결과를 보면 지금까지도 그 개선안이 실천되었다고 보기에는 미흡한 점이 있다고 생각된다.

4.2 주제 분석

2001년에서 2020년까지 발표된 교양영어 교육 관련 논문들의 영문초록을 대상으로 LDA 분석을 실시하였다. LDA 분석은 주제 수에 따라 키워드 수집 결과가 달라지며, 주제 수를 결정하고 추출된 키워드에 대해 주제명을 부여하는 것은 연구자의 역할이다. 주제 수 결정은 주제수에 따른 주제별 키워드를 보면서 정성적으로 판단하는 방법과 주제 수에 따른 실루엣 계수를 측정하여 정량적으로 판단하는 방법이 있다(하근희, 2021). 실루엣 계수는 LDA 분석에서 주제 개수 설정의 신뢰도를 확인하는 방법으로 주제 그룹(cluster) 간의 차이가 클수록, 즉 주제 분류가 잘 이루어졌을수록 1에 가까운 수치로 나타난다고 알려져 있다. 넷마이너에서는 K-means 방법으로 실루엣 계수를 분석할 수 있다.

주제 수 결정을 위해 주제 개수를 계속 변경해나가면서 LDA 분석을 실시하였다. 추출된 키워드와 실루엣 계수를 확인한 결과, 교양영어 교육의 주제를 10개로 분류하는 것이 가장 적절하다고 판단하였다. 첫째, 주제를 10개로 분류했을 때 선행 연구에서 언급된 주요 주제들이 모두 키워드에 포함되어 있었고 둘째, 주제를 10개로 설정하였

을 때 실루엣 계수가 0.839로 매우 높게 나타났기 때문이다. <표 3>은 주제를 10개로 분류한 LDA 분석 결과를 옮긴 것으로 각 주제별 키워드와 각 주제에 해당되는 논문수를 보여준다. 분석 대상이 영문 초록이었으므로 키워드는 모두 영어로 추출되었고, 주제명은 숫자로만 구분되어나타난다. 주제별로 키워드 1~5까지 화면에 제시되며,키워드의 중요도는 키워드1이 가장 높고 키워드5로 갈수록 낮아진다.

LDA 토픽 모델링 분석결과를 바탕으로 각 주제별 제목을 부여하는 단계를 진행하였다. 이를 위해 각 주제별로 추출된 키워드를 중요하게 고려하였고, 해당 범주로 분류된 논문들의 제목, 저자 키워드, 초록도 참고하였다. 필요한 경우 각 주제별로 주제 가능성(Topic probability)이 높은 연구 논문을 직접 확인하기도 하였다. 이와 같은 단계를 거쳐 만들어진 주제1~10의 주제명과 주제별 키워드는 <표 4>와 같다. 아울러 <표 3>에서 제시한 논문수를 기준으로 각 주제가 전체 데이터에서 차지하는 비율을 백분율로 표시하였다.

교양영어 교육의 주제를 하나씩 살펴보면 다음과 같다. 주제1, '교육과정 및 수준별 수업'의 키워드로 추출된 단어는 'course', 'questionnaire', 'level-differentiated', 'program', 'professors'였다. 이 주제에 해당되는 논문은 교양영어 교육과정을 분석하거나 수준별 교육과정에 대한 효과, 교육과정에 대한 학습자 설문 결과를 보고한 논문들이었다. 주제1의 논문 수는 56편으로 전체 토픽 중에서 가장 높은 비율인 18.67%를 차지하였다.

주제2, '학습자 정의적 요인'의 키워드로 추출된 단어는 'motivation', 'difference', 'questionnaire', 'reading', 'anxiety'였다. 학습 동기나 학습 불안, 자기 효능감과 같

번호	주제 주제	주제별 키워드	논문 수	백분율
1	교육과정 및 수준별 수업	교육과정 분석, 수준별 프로그램, 레벨 테스트	56	18.67%
2	학습자 정의적 요인	학습 동기, 탈동기화, 학습 불안, 자기 효능감	28	9.33%
3	평가 및 학습 전략	TOEIC, 분반평가, 시험전략, 독해전략	27	9.00%
4	교수자 요인	원어민 교사, 비원어민 교사, TETE	32	10.67%
5	어휘 • 문법학습 및 쓰기활동	어휘학습, 문법기반 교수, 영어쓰기, 오류 분석	27	9.00%
6	특수목적영어 프로그램	ESP, 직업, 전공, 튜터링 프로그램, EAP	25	8.33%
7	교수학습 방법	Flipped learning, Problem-based, 프로젝트 학습	21	7.00%
8	웹 기반 수업	온라인, 인터넷, 사이버, TED, 소셜 미디어	20	6.67%
9	학습자 중심 교육	수업 활동, 소그룹 활동, 협동학습, 학습자 만족도	35	11.67%
10	교재분석 및 기타	교재분석, 교양 교육, 역량, 문화, General education	29	9.67%
합계			300	100.00%

⟨표 4⟩ 교양영어 교육 주제와 관련 키워드

은 학습자의 정의적 요소와 교양영어 학습 간의 관련성을 다룬 연구가 이 주제로 분류되었다. 선행연구에서는 '학 습동기/전략', 혹은 '학습심리'로 명명하고 있으나 본 연 구에서는 학습과 관련한 학습자의 심리를 좀 더 포괄적으 로 표현하고자 '학습자 정의적 요인'으로 명명하였다. 이 주제에 해당하는 논문은 28편으로 전체의 9.33%에 해당 하였다.

주제3, '평가 및 학습 전략'의 키워드로 추출된 단어는 'test', 'level-differentiated', 'strategy', 'score', 'TOEIC' 이었다. 추출된 키워드에서 드러나듯이 이 주제에는 시험과 관련된 연구들이 많이 포함되었으며 TOEIC도 이 주제로 분류되었다. 교양영어 교육에서 실용영어의 중요성이 강조되고, 영어졸업인증제를 실시하는 많은 대학에서 TOEIC 성적을 반영하는 등 교양영어 수업에서 관련 이슈가 중요하게 다루어졌기 때문인 것으로 풀이된다. 주제 3에 해당되는 논문은 27편이었으며 전체의 10%에 해당되었다.

주제4, '교수자 요인'의 키워드로 추출된 단어는 'questionnaire', 'NEST', 'interview', 'experimental group', 'control group'이었다. 이 주제는 교수자와 관련된 연구들로, 특히 원어민 교수자(NEST, Native English Speaking Teacher)의 수업 효과와 관련된 연구가 다수 포함되어 있었다. 이외에도 원어민 교수자와 내국인 교수자의 수업 효과를 비교한 연구, 영어로 진행하는 영어 수업(TETE, Teaching English Through English)과 관련된 연구들도이 범주에 포함되었다. 주제4에 해당되는 논문은 총 32편으로 전체의 10.67%였으며, 세번째로 많은 연구가 이루어진 주제였다.

주제5, '어휘·문법학습 및 쓰기활동'의 키워드로 추출 된 명사들은 'vocabulary', 'grammar instruction', 'test', 'error', 'type'이었다. 이 주제는 명시적 영어 학습과 관련 된 주제로 어휘 학습이나 문법 학습, 제2언어습득을 위한 쓰기 활동, 즉 'writing to learn' (Manchón, 2011)과 관련 된 연구들이었고, 협력 쓰기(collaborative writing)를 연구 한 논문들도 이 범주로 분류되었다. 이 주제로 분류된 논문은 총 27편이었으며 전체의 9%에 해당되었다.

주제6, '특수목적영어 프로그램'의 키워드 1~5는 'program', 'curriculum', 'tutee', 'liberal arts', 'need'였다. 표에는 제시되지 않았지만 주제6의 키워드 6~10은 'presentation', 'job', 'ESP', 'major', 'subject'였다. 키워드1~10을 보면알 수 있듯이 이 주제는 취업 역량 혹은 전공 역량의 강화를위한 교양영어 수업인 것으로 나타났다. 교양영어 수업과 관련한 학생들의 현실적인 기대와 그에 대한 대학의 노력이 이 주제로 나타난 것이라 보인다. 여기서 주제6의 주제명, '특수목적영어'은 취업역량 강화, 전공 영어 학습역량강화, 프리젠테이션 역량 강화 등 특수한 목적을 위해실시되는 교양영어 수업을 의미한다. 이 주제에 해당하는 논문은 총 25편이었으며 전체의 8.33%에 해당하였다.

주제7, '교수학습방법'의 키워드로는 'flipped learning', 'blended learning', 'classroom', 'interests', 'movie'가 수집되었다. 이 범주에는 특히 flipped learning 관련 연구가많이 포함되어 있었다. Flipped learning은 Baker(2000)에의해 처음 제안된 수업 방법으로, 플립드 러닝, 플립러닝, 거꾸로 교실, 거꾸로 수업, 역 진행 수업 등 다양한 용어로사용된다. 주로 교수자가 사전에 준비한 수업 영상과 자료를 학생이 수업시간 전에 미리 보고, 교실 수업시간에는

학생들이 상호작용을 하면서 배우거나 심화학습을 하도록 하는 수업방식을 말한다. 두 번째 키워드인 Blended learning은 온라인과 교실수업을 결합하는 방식을 말하며, 'hybrid learning', 'technology-mediated instruction', 'web-enhanced instruction' 등으로 언급되기도 한다. 이처럼 다양한 교수법이 교양영어 교육의 주제 중 한 가지로 드러난 것은 교양영어 교실에서 혁신 수업 기술이 많이도입되고 연구되고 있음을 보여주는 예라 할 수 있겠다. 주제7로 분류된 논문 수는 21편으로 전체의 7%에 해당하였다.

주제8, '웹 기반 수업'의 키워드 1~5는 'course', 'content', 'professors', 'online class', 'online'이었다. 이 외에도 주제8의 키워드로 추출된 단어로 'video', 'TedED', 'VR', 'web-based', 'LMS (Learning Management System)', 'application' 등이 있었다. 제시한 키워드에서도 드러나듯이, 온라인 학습과 온라인 학습 관리, 인터넷, 비디오, 앱등을 이용한 수업에 대해 연구한 논문들이 이 범주에 해당되었다. 주제8에 해당되는 논문은 총 20편으로 전체의 6.67%에 해당되었다. 이 주제에 해당하는 논문들은 1편을 제외하고는 모두 2010년 이후에 출판된 논문들이었으며 이는 다양한 온라인 학습 자료와 도구, 그리고 최근 몇 년 동안에 활발히 이루어지고 있는 온라인 학습 환경과 관련이 있을 것으로 생각된다.

주제9, '학습자 중심 교육'의 키워드로는 'activity', 'questionnaire', 'environment', 'classroom', 'role'이 추출되었다. 이외에도 이 주제의 키워드로 '학습자 만족도', '조별 활동', '소그룹 활동', 'cooperative', 'learners' satisfaction', 'learner-centered' 등의 어휘들이 등장하였다. 주제명에서 '학습자 중심'이란 표현은 학생 활동이 중심이 되는 수업 이라는 의미와 함께 수업활동에 대한 학습자의 만족, 학습자 요구 중심의 교육이라는 의미를 강조하기 위해 사용되었다. 이 주제에 해당되는 연구는 총 35편으로 전체의 11.67%였으며, '교육과정 및 수준별 수업'에 이어 두 번째로 많이 연구된 주제였다.

주제10, '교재분석 및 기타'의 키워드는 'textbooks', 'competence', 'knowledge', 'ability', 'development'였다. 가장 중요한 키워드인 'textbooks' 이외에 다양한 스펙트럼을 가진 키워드가 이 주제에 포함된 것을 알 수 있다. 이에 주제10에 해당되는 논문들의 서지정보를 확인한 결과, 가장 큰 비중을 차지하는 키워드인 '교재'를 비롯하여

'역량' '교양교육', '교육현황', '연구동향', '비교과', '융합', '문화' '글로벌 역량' 등의 다양한 키워드들이 수집된 것을 확인할 수 있었다. 이와 더불어 '교양교육으로서의 교양영어', '문화 교육', '문학을 이용한 영어교육', '교양영어 교육의 개선 방안' 등 교양영어를 거시적 안목에서 분석하고 연구한 연구들도 이 범주에 포함되어 있었다. 이에 주제10의 주제명을 '교재 분석 및 기타'로 정하였다. 주제10에 해당되는 연구는 총 29편으로 전체의 9.67%에 해당하였다.

4.3 연구 동향 분석

교양영어 교육 주제들의 연구동향을 알아보고자 주제별 논문 발행 수를 시기별로 분석하였다. 앞서 연도별 논문 발행 수에서도 살펴보았듯이 2000년도에 비해 2010년도에는 각 주제별 논문 수가 훨씬 증가하였으며, 주제별로 논문이 폭넓게 출판되는 경향을 확인할 수 있다. 즉최근 20년간 대학 교양영어 교육 관련 연구가 양적으로증가했을 뿐 아니라 연구 주제 면에서도 훨씬 다양화되고 있음을 알 수 있었다. <표 5>는 주제별 논문 발행 수를 5년 단위로 분석해 본 수치이다. 시기는 5년 단위로 표시하였으며 숫자는 시기별 논문 수, 백분율은 해당 시기에서 그 주제가 차지하는 비율을 의미한다.

[그림 2]는 5년 단위 시기별로 각 주제 비중의 변화를 막대 그래프로 나타낸 것이다. 5년 단위 시기별 발행 논문수에 차이가 있음에도 시기별 막대의 길이가 같은 것은 5년 단위 전체 논문수를 100%로 보았을 때, 각 주제(T1, T2, ... T10)가 그 시기에 몇 %의 비중을 차지하는지를 파악하고자 함이다. 기술의 편의상 2001년에서 2005년까지를 [시기1], 2006년에서 2010년까지를 [시기2], 2011년에서 2015년까지를 [시기3], 2016년에서 2020년까지를 [시기4]로 명하고 논의를 이어가고자 한다.

[시기1]~[시기3] 기간 동안 가장 많은 연구가 이루어진 주제는 주제1, '교육과정과 수준별 수업'이었다. 그러나 [시기4]가 되면 주제1의 연구가 현저히 줄어들어드는 것을 확인할 수 있다. [시기4]에서 가장 많은 연구가 이루어진 주제는 주제9, '학습자 중심교육'이었다. 주제9, '학습자 중심 교육'은 [시기2],[시기3]에 비해 [시기4]에 두 배이상 논문 수가 증가한 것으로 나타났다. 주제2, '학습자정의적 요인'과 주제5, '어휘·문법 학습 및 쓰기활동'은

번호	주제 명	2001-2005	2006-2010	2011-2015	2016-2020
주제1	교육과정 및 수준별 수업	6(30.00%)	10(23.80%)	23(27.05%)	17(11.11%)
주제2	학습자 정의적 요인	0(0.00%)	3(7.14%)	4(4.70%)	21(13.73)%
주제3	평가 및 학습 전략	2(10.00%)	2(4.76%)	13(15.29%)	10(6.54%)
주제4	교수자 요인	2(10.00%)	5(11.90%)	14(16.47%)	11(7.19%)
주제5	어휘 • 문법학습 및 쓰기활동	0(0.00%)	3(7.14%)	6(7.06%)	18(11.76%)
주제6	특수목적영어 프로그램	1(5.00%)	4(9.52%)	5(5.88%)	15(9.80%)
주제7	교수학습 방법	3(15.00%)	3(7.14%)	3(3.53%)	12(7.84)%
주제8	웹 기반 수업	1(5.00%)	4(9.52%)	5(5.88%)	10(6.54%)
주제9	학습자 중심 교육	2(10.00%)	3(7.14%)	6(7.06%)	24(15.69%)
주제10	교재 분석 및 기타	3(15.00%)	5(11.90%)	6(7.06%)	15(9.80%)
합계		20(100.00%)	42(100.00%)	85(100.00%)	153(100.00%)

〈표 5〉 주제별 논문 수 및 백분율(2001~2020)



[그림 2] 주제별 논문 발행 수 5년 단위 추이

[시기1]에서는 등장하지 않다가 [시기2]부터 등장한다는 공통점이 있다. 하지만 주제5, '어휘·문법 학습 및 쓰기활동'은 [시기2] ~ [시기4]까지 논문 수 비중의 변화가 크게 없는데 반해, 주제2 '학습자 정의적 요인'은 [시기4]에서 비중이 크게 증가한 것을 알 수 있다.

지금까지 지난 20년간 각 주제의 논문 수 추이를 간략 하게나마 확인해보았다. 이제부터 논문 수 추이를 바탕으로 주제별 향후 전망을 알아보고자 한다. 주제별 연구 전망을 알아보기 위한 방법으로 시계열 분석을 활용하였다. 시계열분석이란 시간의 흐름에 따라 독립변수 X에 대한 Y의 변화량을 예측하는 방법으로, 시계열 분석의

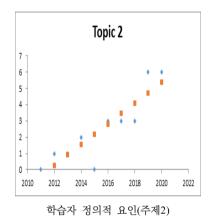
가장 큰 장점은 과거를 설명하는 동시에 미래를 예측할 수 있다는 점이다(Ostrom, 1990).

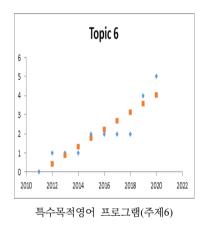
분석 기간 전체를 대상으로 시계열 분석을 실시한 결과, 2001년에서 2020년까지의 20년간은 논문 수 자체가크게 증가하였으므로, 10개의 주제 모두 유의미한 증가가 있는 것으로 나타났다. 이에 논문 수가 큰 폭으로 증가한시기인 2001년에서 2010년을 제외하고, 2011년에서 2020년 사이 연간 주제별 논문 수 데이터로 시계열 분석을 실시하였다. <표 6>은 각 주제별 시계열 분석 결과를 나타낸다.

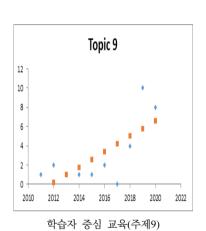
<표 6>에서 나타나는 바와 같이 시계열 분석 결과 3가

₹₩	6)	주제볔	시계연	보선	결과

주제	유의도	분석 결과	주제	유의도	분석 결과
1	0.725	-	6	0.001	상승 주제
2	0.001	상승 주제	7	0.194	-
3	0.761	-	8	0.103	-
4	0.927	-	9	0.019	상승 주제
5	0.211	-	10	0.073	-







[그림 3] 상승 주제 시계열 분석결과

지 주제가 10년의 기간 동안 유의미한 논문 수 증가가 있었던 것으로 나타났다. 여기서 논문 수가 유의미하게 증가했다는 의미는 향후 관련 연구가 늘어날 가능성이 높다는 뜻으로 해석될 수 있다. 분석 결과 상승 주제로 나타난 주제는 주제2, '학습자 정의적 요인', 주제6, '특수 목적 영어 프로그램', 주제9, '학습자 중심교육' 이었다. 이 3가지를 제외한 나머지 7개 주제는 상승 혹은 하강의 특별한 경향성을 띠지는 않는 것으로 나타났다.

[그림 3]은 상승 주제로 분류된 주제들의 시계열 분석결 과 그래프이다. X축은 년도를 나타내며 시기는 2011년에 서 2020년까지이다. Y축은 논문 수를 나타내며 주제별로 최대 숫자가 다르게 나타난다. 이는 주제별로 논문 발행 수 최대치가 다르므로, Y축의 숫자가 주제에 따라 0~7(주 제2), 0~6(주제6), 0~12(주제9)로 각각 다르게 나타난 것 이다. 그림에서 일직선으로 보이는 작은 네모로 이어진 선은 분석을 위한 예측치이며, 일직선과 겹치기도 하면서 다소 산발적으로 나타나는 마름모 표시가 실제 출판된 논문 수이다.

상승 주제를 보면, '학습자 정의적 요인'과 '학습자 중 심 교육'은 둘 다 학습자와 직접적으로 관련된 주제임을 알 수 있다. '특수목적영어 프로그램'은 학습자의 직업

역량, 전공영어 역량 강화와 같은 학습자의 기대와 요구에 부응하고자 설계된 교양교육 프로그램이므로 역시 학습 자의 니즈와 관련이 있는 주제라 하겠다. 이런 맥락에서 볼 때 상승 주제로 파악된 세 가지 주제가 넓은 의미에서 모두 학습자와 관련이 있는 주제라고 볼 수 있을 것이다. 상승 주제로 파악된 주제가 모두 학습자의 입장을 고려한 연구주제라는 점은 교양영어 수업 현장과 관련 연구에 있어서 시사하는 바가 적지 않다고 생각된다.

즉, 교양영어 교육과 연구는 학습자의 필요에 부응하 고, 학습자 중심의 수업활동을 실시하며, 더불어 학습자의 동기나 불안과 같은 정의적 영역을 살피는 방향으로 진행 될 것으로 분석결과 드러난 것이다. 1960년대 교수자 강 독 중심의 교양영어 수업을 하던 단계를 지나 회화 중심 실용영어 수업, 통합적 영어의사소통역량 개발이라는 교 육과정의 변화를 겪어온 교양영어 교육은 이제 2020년대 에 이르러 학습자 중심 교양영어 수업으로 전환되고 있는 것처럼 보인다. 학습자 요구, 학습자 내면, 학습자 활동 중심의 수업과 연구가 향후에도 비중이 높아질 가능성이 많다는 것을 이 분석결과가 알려주고 있기 때문이다.

이제 교양영어 교육은 학습자의 동기를 끌어올리고, 학습 불안과 같은 내면을 살피면서, 학습자 집단이 필요로 하는 현실적 목표에 부응하고, 학습자가 스스로 배우는, 학습자 요소와 관련한 주제에 주목할 필요가 있다고 데이 터가 우리에게 알려주고 있는 것이다.

5. 결론

본 연구는 2001년부터 2020년까지 대학 교양영어 관련 연구의 연구동향을 분석함에 있어, 총 313편의 논문을 대상으로 연도별 논문 발행 수, 연구 주제, 주제별 연구동향을 살펴보았다. 분석 결과, 2001년에서 2020년까지 대학 교양영어 관련 연구는 양적으로 큰 폭의 증가가 있었던 것으로 나타났다. 분석기간을 5년 단위로 나누었을때, 2001년~2005년 사이에 발표된 논문은 26편, 2006년~2010년 사이에는 44편, 2011년~2015년 사이에는 89편, 2016년~2020년 사이에는 154편으로 양적으로 꾸준한증가세를 보였으며 시기가 진행될수록 연구 주제도 다양해졌음을 알 수 있었다.

둘째, 동 기간 교양영어 교육의 연구 주제를 LDA 토픽모델링 기법으로 분석해보았다. 주제 수는 추출된 키워드와 실루엣 계수라는 토픽 분류 신뢰도 측정 방법을 근거로하여 10개의 주제로 분류되었다. 주제명은 '교육과정 및수준별 수업(주제1)', '학습자 정의적 요인(주제2)', '평가및 학습 전략(주제3)', '교수자 요인(주제4)', '어휘・문법학습 및 쓰기 활동(주제5)', '특수목적영어 프로그램(주제6)', '교수학습 방법(주제7)', '웹 기반 수업(주제8)', '학습자중심 교육(주제9)', '교재분석 및 기타(주제10)'였다. 20년의 기간 동안 가장 많이 연구된 주제는 '교육과정및 수준별 수업(주제1)'이었으며 2016~2020년, 최근 5년간 가장 많이 연구된 주제는 '학습자 중심 교육(주제9)'이었다.

세 번째, 연구 주제별 전망 분석을 위해 10개의 연구 주제를 대상으로 2011년에서 2020년의 기간을 설정하여 시계열 분석을 실시하였다. 시계열 분석 결과, '학습자 정의적 요인(주제2)', '특수목적 영어 프로그램(주제6)', '학습자 중심 교육(주제9)', 세 가지가 상승 주제인 것으로 드러났다. 상승 주제로 전망된 연구 주제가 모두 학습자의 학습 동기, 학습자의 니즈, 학습 활동과 관련된 주제라는 점에 비추어볼 때 향후 교양영어 교육 연구에 있어서 학습 자 요소가 주요한 연구 트렌드가 될 것으로 전망되었다.

본 연구는 LDA 토픽 모델링 방식으로 주제 분석과 주제별 동향 분석을 실시함으로써 선행 연구에 비해 좀 더 객관적이고 유용한 분석결과를 도출하였다는데 의의 가 있다고 본다. 본 연구를 통해 LDA 토픽모델링 분석 방법을 활용해 본 결과, 첫째, 각 논문에 포함된 모든 주제 를 주제 분석에 반영한다는 점, 둘째, 모든 데이터를 분석 한 후 주제를 추출해내고, 각 문서를 주제별로 분류해낸다 는 점에서 분석결과에 대한 신뢰도가 높아지는 장점이 있음을 알 수 있었다. 하지만 LDA 분석에서도 전처리 단계, 주제 수를 결정하는 단계, 주제명을 부여하는 단계 에서 연구자의 역할과 판단이 요구된다는 점에서 이 분석 방법도 온전히 데이터 기반 분석방법이라고 보기에는 어 려운 측면이 있음을 알 수 있었다. 하지만 데이터 기반으 로 분석하는 LDA 토픽 모델링 기법이 연구자의 판단에 의한 주제 분석 방법보다는 크게 진일보한 연구 방법이라 는 것은 분명하다고 본다.

본 연구를 마무리하면서 두 가지 점을 제안하고자 한다. 첫째는 학회에서 학술 논문 접수 시 서지 정보(제목, 저자 키워드, 초록)를 우리말과 영어, 두 언어로 제출하도록 지정할 필요가 있다는 점이다. 본 연구에서는 영문 초록이 없는 13편의 논문들을 토픽모델링 분석에서 제외할 수 밖에 없었다. 향후 비정형 텍스트 데이터를 이용한분석은 점차 활발해질 것으로 예상되며, 연구자들은 연구목적에 따라 영어 혹은 우리말 서지 정보를 선택하여 연구할 수 있으므로, 소중한 연구 결과가 제외되지 않기 위해서는 영어와 우리말 서지정보를 모두 수집할 필요가 있다고 생각된다.

둘째, 교양영어가 영어교육의 한 분야로 자리매김할 필요가 있음을 강조하고자 한다. 본 연구에서 교양영어 저자 분석을 실시해본 결과, 교양영어를 전문적으로 연구 하는 연구자 수가 극히 적다는 것을 알 수 있었다. 대학의 교양영어 교육은 전체 학과의 학생들을 대상으로 한다는 점, 대학교육의 근간이 되는 교양교육의 핵심적인 부분인 점, 그리고 무엇보다도 대학교육을 받고 있는 대부분의 학생들은 영어사용능력 향상에 대한 큰 기대와 갈망이 있다는 점에서 교양영어 교육의 중요성이 재고되고 관련 연구가 더 활발히 진행될 필요가 있음을 제안하고자 한다.

참고문헌

- 김상수(2009). "대학 교양영어교육의 효율적 운영방안에 관한 연구", 영남대학교 미출간 석사학위논문.
- 김성혜, 임자연(2013). "대학 교양영어 프로그램의 운영 현황", *현대영어교육* 14(2), 263-290.
- 김종보(1997). "의사소통능력 향상을 위한 교수방법 연구 전문 대학 교양영어를 중심으로", *영어교육연구* 6(1), 147~163.
- 박은희(2021). "텍스트 마이닝을 활용한 영어 쓰기교육 연구동향 분석", *응용언어학* 37(3), 95-122.
- 방영주(2004). "영어능력인증제, 능력별 교과과정, 그리고 대학 교양영어교육: 학생과 교수자의 인식에 관한 고찰", *영어어 문교육* 9(2), 193-211.
- 방영주(2014). "효율적인 대학 교양영어교육의 방향 및 제언", *현대영어교육* 15(4), 221-245.
- 성명희, 장형지, 김현진(2007). "교양영어교육에 관한 학생 요구 조사: 보건 계열 특성화 대학의 사례를 중심으로", *현대영어 교육* 8(3), 318-340.
- 송미정, 박용예(2004). "대학교양영어교육의 통합적 접근", *영어 교육* 59(2), 179-211.
- 원영국, 김영우(2021). "토픽 모델링을 활용한 한국 영어교육 학술지에 나타난 연구동향 분석", *한국콘텐츠학회* 21(4), 50-59.
- 이연숙(2014). "대학 영어 프로그램에 대한 선행연구 분석과 앞으로의 연구방향", *언어과학연구* 70, 353-384.
- 이효웅(1996). "교양영어교육의 문제점과 개선 방향", *영어교육 연구* 5(1), 105-129.
- 임수민, 윤회정, 방담이(2020). "주제어 네트워크 분석을 통한 교양교육연구 동향 분석", 교양교육연구 14(1), 11-32.
- 조성준(2019). *세상을 읽는 새로운 언어, 빅데이터,* 파주: 21세기

북스.

- 조정순(2002). "대학 교양영어 교육의 현황과 방향성 탐구", *영어 교육* 57(2), 365-394.
- 최원경(2020). "면대면 대 비대면 강의 만족도 비교: 코로나19 사태에서의 대학원 교양영어 수업 사례 연구", *영어교과교* 육 19(4), 223-245.
- 하근희(2021). "LDA를 활용한 초등국어교육 연구 동향분석-KCI 등록 논문을 중심으로", *한국초등국어교육* 70, 295-318.
- 하명애(2017). "대학 교양영어에 대한 연구 동향 분석", *언어연구* 33(3), 453-476.
- Baker, J. W.(2000). "The "classroom flip": Using web course management tools to become the guide on the side", In J. A. Chambers (Ed.), Selected papers from the 11th International Conference on College Teaching and Learning (pp. 9-17), Jacksonville, Florida.
- Blei, D.(2012). "Probabilistic top models", *Communications of the ACM* 55(4), 77-84.
- Blei, D., Ng, A., & Jordan, M.(2003). "Latent Dirichlet allocation", Journal of Machine Learning Research 3, 993-1022.
- Maier, D., Waldherr, A., Miltner, P., Weidemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Haussler, T., schmid-Petri, H., & Adam, S.(2018). "Applying LDA topic modeling in communication research: Toward a valid and reliable methodology", Communication Methods and Measures 12(2-3), 93-118.
- Manchón, R.(2011). *Learning-to-Write and Writing-to-Learn in an Additional Language*, Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ostrom, C. W.(1990). *Time Series Analysis: Regression techniques*, Thousand Oaks, California: Sage Publications.

Research Trends in College English Education in Korea

- A Topic Analysis Using LDA Topic Modeling

Park, Eunhee

Lecturer, Pukyong National University

Abstract

This study investigated the research trends of college English education in Korea from 2001 to 2020. The data was collected using a Biblio data collector and a total of 313 papers were analyzed. For research purposes, the data were analyzed using frequency analysis, LDA (Latent Dirichlet Allocation), and time series analysis. The summary of the findings is as follows: In the first instance, the number of research papers regarding college English education has increased significantly in quantity for 20 years. Secondly, in analyzing the topics of the chosen papers, a total of 10 topics in college English education were found. The topics were "curriculum and level-differentiated programs (T1)", "learners' affective factors (T2)", "assessment and learning strategies (T3)", "teachers' factors (T4)", "English vocabulary, grammar and writing (T5)", "English for specific purposes (T6)", "teaching and learning methods (T7)", "web-based learning (T8)", "learner-centered education (T9)", and "textbook analysis etc. (T10)." Among these topics, the three that were identified as topics increasing in popularity were "learners' affective factors (T2)", "English for specific purposes (T6)", and "learner-centered education (T9)." The topics increasing in popularity shared one key characteristic: the topics were related to learners' factors such as the learners' motivation, the learners' goals, and the learners' activities in class. This study is meaningful in that it collected a wide range of data related to college English education in Korea and produced more reliable results by using big data-based LDA topic modeling techniques.

Key Words: college English education, university English, research trends, topic analysis, topic modeling, LDA (Latent Dirichlet Allocation)