



최근 10 년간(2014~2023) 《영어학》 연구 동향 분석: 계량정보학적 접근*

남재우 (건국대학교) · 장인철 (서울대학교)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: April 6, 2024

Revised: April 17, 2024

Accepted: April 18, 2024

Nam, Jaewoo
Professor, Department of Library
and Information Science
Konkuk University
268 Chungwon-daero Chungju-si
Chungcheongbuk-do, Korea
Tel: 043) 840-3369
Email: lisnam@kku.ac.kr

Jang, In Chull
Professor, Department of English
Language Education
Seoul National University
1 Gwanak-ro Gwanak-gu
Seoul, Korea
Tel: 02) 880-7676
Email: icjang@snu.ac.kr

* This paper was supported by
Konkuk University in 2023

ABSTRACT

Nam, Jaewoo and In Chull Jang. 2024. An analysis of research trends of Korea Journal of English Language and Linguistics over the past decade (2014~2023): An informetric approach. *Korea Journal of English Language and Linguistics* 24, 348-374.

Using text mining methods, this study examines research trends in papers published in Korea Journal of English Language and Linguistics over the past decade (2014-2023). A total of 456 abstracts were collected and analyzed using frequency analysis, Phi coefficient, TF-IDF, and LDA topic modeling. The findings are as follows. First, applied linguistics and English education are dominant fields in the journal, likely reflecting the status of English as a second or foreign language in South Korea. Second, within applied linguistics and English education, writing and vocabulary emerge as major research themes, particularly academic writing in the English for Academic Purposes (EAP) context. Third, with general linguistics, syntax exhibits significant research output both in quantity and specificity of topics. Fourth, emerging research areas include corpus linguistics, psycholinguistics, computational linguistics, and sociolinguistics. Fifth, research topics display diversity and evolution, driven by the discovery of new English-related phenomena. In light of these findings, this study underscores the ongoing necessity for research in English linguistics and language, emphasizing the importance of staying attuned to social changes and disciplinary identities

KEYWORDS

research trends, text mining, TF-IDF, topic modeling, phi coefficient, English linguistics, English language, disciplinary identity, informetrics

1. 서론

학문의 성취와 추이를 체계적으로 분석하고 서지에 관해 연구하는 것은 서지학의 전통적인 분야이다. 최근 서지학은 현대의 정보통신기술과 접목해 학술정보의 흐름과 이용을 분석하는 계량정보학(informetrics) 분야로 발전하였다. 계량정보학은 수학적, 통계적 방법을 사용하여 특정 분야의 문헌을 체계적으로 분석하여 연구 분야의 주제 및 연구 동향 등 연구 결과를 종합하는 접근 방식이다(Pritchard 1969). 이는 많은 양의 텍스트 데이터를 통계적으로 분석하여 결과를 제시해주므로, 연구동향을 빠르게 파악할 수 있는 연구방법론이다. 또한, 주관적인 판단이나 편견을 최소화하고, 연구 결과의 재현과 검증이 가능한 상대적으로 객관적인 문헌 연구 방법이기도 하다(Lawani 1981). 따라서 본 연구는 계량정보학적 관점에서 영어학 분야의 대표적 국내 학술지인 《영어학》(Korean Journal of English Language and Linguistics)의 연구주제와 동향을 분석하여 학술지의 학문적 기여를 파악하고, 이를 바탕으로 학술지의 학문적 발전을 위한 시사점을 제안하고자 한다.

《영어학》은 한국영어학회에서 발간하는 학술지로 2000 년 학회가 창립된 후, 2001 년 2 월에 창간호를 발간하였다.¹ 2004 년 등재후보학술지, 2007 년 등재학술지, 2019 년 우수등재학술지로 선정되었고, 2020 년 SCOPUS 에 등재되었다. 《영어학》은 “추상적인 언어원리의 추구보다는 영어 자체에 대한 관심과 영어의 실증적인 면에 대한 연구를 목표”²로 하고 있으며, “영어사, 영문법, 음성학, 음운론, 통사론, 의미론, 화용론, 담화분석, 응용언어학(e.g., 제 2 언어습득, 영어교육)을 포함한 영어학의 모든 분야(all fields of English linguistics, including English history, grammar, phonetics, phonology, syntax, semantics, pragmatics, discourse analyses, and even applied linguistics (e.g., second language acquisition or English language teaching))”³에 관한 연구 논문을 출간한다.

《영어학》은 한국에서 발행하는 언어학 분야의 우수등재지이자 SCOPUS 에 등재되어 있다. 한국연구재단 우수학술지 등재와 SCOPUS 등재가 학술지의 권위와 질을 모든 측면에서 보장하지는 않지만 해당 기관의 심사를 거쳐 획득하는 만큼 학술지의 국내외적 위상을 나타낸다는 점에서 《영어학》의 학술지로서 중요성을 보여준다. 따라서, 국내 영어학 분야의 대표적 학술지인 《영어학》의 연구동향을 파악하는 것은 한국 영어학 연구를 조망하고 발전 방향을 모색하는 작업이라고 할 수 있다. 이를 통해 영어학 세부 연구 영역의 양상과 흐름을 파악하여 학문의 정체성 확립에 기여하고, 영어학 연구자들에게 향후 연구 방향을 제시하는 데 도움이 될 것이다. 또한, 다른 학술지와 달리 영어학이라는 학문 분야에 친착하는 《영어학》 학술지의 연구 범위를 고려했을 때, 해당 학술지의 연구 동향 분석은 언어학, 응용언어학, 영어교육학 등 인접 학문과의 관계를 파악하는 데도 중요하다.

본 논문은 《영어학》에 게재된 10 년간(2014~2023)의 논문을 분석 대상으로 설정하였다. 《영어학》은 2001 년부터 발간되었지만, 모든 기간을 포함시키면 최신 연구동향을 파악하기 어렵고, 반면 최근 5 년의 논문을 대상으로 하면 분석 데이터가 부족해 정확한 결과를 도출하기

¹ 한국영어학회 홈페이지의 학회 역사(<http://www.kasell.or.kr/20>)에 따르면, 2000 년 8 월 23 일 서강대학교에서 창립총회를 했으나, 학술지 홈페이지(<http://journal.kasell.or.kr>)의 “Aims and Scope”에 기술된 설명에 따르면, 2000 년 5 월 8 일에 창립되었다.

² <http://www.kasell.or.kr/20>

³ <http://journal.kasell.or.kr>

힘들다. 따라서 지난 10 년은 비교적 최근의 영어학 분야의 발전 과정을 탐색할 수 있으며, 동시에 충분한 양의 데이터를 확보하여 연구의 신뢰성을 높일 수 있는 적절한 기간으로 판단하였다.

2. 이론적 배경

2.1 텍스트 마이닝

텍스트 마이닝(Text Mining)은 데이터 마이닝(Data Mining)의 한 분야로, 대규모 텍스트 데이터에서 유용한 정보를 추출하는 분석방법이다(Feldman and Dagan 1995). 대규모 텍스트 데이터에서 의미 있는 패턴과 추세를 파악하고 잠재적인 지식을 추출하는 것은 다양한 분야에서 중요한 작업이다(Anne and Stephen 2007). 과거 많은 연구자들은 수동 코딩이나 담화분석 방법으로 텍스트 데이터를 분석하는 질적 연구방법을 사용했으나, 대규모 텍스트 데이터 분석에는 한계가 있었다(Kobayashi et al. 2018). 이에 자동화된 텍스트 데이터 분석방법을 탐색하였고(Wiedemann 2013), 이 과정에서 텍스트 마이닝은 강력한 도구로 활용되었다(Eroğlu and Seçkiner 2016). 현재 텍스트 마이닝은 기계학습, 정보검색, 정보추출 등 다양한 분야에서 활용되며, 학술 연구에서는 연구 동향과 주제를 발견하는 데 이용되고 있다(Aurora et al. 2007).

텍스트 데이터는 주로 인간의 언어인 자연어로 구성되어 있으므로, 컴퓨터가 읽을 수 있도록 가공하는 자연어 처리(natural language process, NLP)를 거친 후, 데이터에서 의미 있는 정보나 패턴을 발견하는 것을 목표로 한다. 데이터는 코퍼스로 구축되며, 이를 통해 텍스트를 수치 정보로 변환하고 단어 빈도 분포를 분석한다(Sebastiani 2002). 텍스트 마이닝은 문서를 하나의 단어 집합으로 취급하는 'Bag of words(BoW)'를 중요 개념으로 사용한다(Harris 1954). 이는 단어의 순서나 문맥을 고려하지 않고 단어의 등장 여부에 주목하여 텍스트를 분석한다는 의미이다(Lee et al. 2010).

2.2 TF-IDF

TF-IDF(Term Frequency-Inverse Document Frequency)는 특정 문서 내에서 단어의 상대적인 중요성을 계산하는 데 사용되는 통계적인 척도이다(Aizawa 2003, Das et al. 2023, Zhang et al. 2011). 이 수치가 높은 단어는 해당 문서에서 상대적으로 중요한 단어로 간주된다(Kim et al. 2017).

TF(Term Frequency)는 특정 단어가 문서 내에서 얼마나 자주 등장하는지를 나타내며, 문서에서 특정 단어가 자주 등장할수록 값이 커진다. IDF(Inverse Document Frequency)는 문서 내에서 특정 단어의 중요성을 나타내는 값으로, 특정 단어의 등장 빈도가 낮을수록 높은 IDF 값을 갖게 된다. TF-IDF 는 TF 와 IDF 를 곱하여 계산되며, 다음과 같은 공식을 갖는다. (단어 w , 문서 d)

$$TF\text{-}IDF(w, d) = TF(w, d) \times IDF(w)$$

특정 문서에 자주 등장하면서 다른 문서에서는 희귀한 단어일수록 높은 TF-IDF 값을 갖게 되며, 이를 통해 특정 단어의 중요성을 빠르게 이해할 수 있다. 이러한 원리로 학술지에서 TF-IDF 값이

큰 단어를 추출하면, 해당 학술지 내 전체 논문에서 중요한 키워드를 식별하여 학술지의 특징과 성격을 파악할 수 있다.

본 연구는 2014 부터 2023 년까지 《영어학》에 게재된 전체 논문을 게재 연도 기준으로 그룹화하여 TF-IDF 분석을 수행하였다. 이는 특정 단어가 해당 연도에서 갖는 중요도를 파악하기 위한 것이다. 예를 들어, 2014 년에 ‘annotation’의 TF-IDF 값이 높았다면, 이 단어는 다른 연도의 논문에서도 등장하지만 2014 년에 특히 다빈도로 등장했거나, 다른 연도의 논문에는 등장하지 않고 2014 년에만 다빈도로 등장한 단어로 볼 수 있다. 따라서, 본 연구는 TF-IDF 값이 큰 단어를 해당 연도의 핵심 키워드로 간주하고, 연도별 핵심 단어의 변화를 통해 연구 동향을 이해하고자 한다.

2.3 LDA 기반 토픽모델링

토픽모델링은 구조화 되지 않은 대규모의 텍스트 데이터에서 일정한 패턴을 발견하여 주제를 찾는 분석방법으로 대표적인 텍스트 마이닝 기법이다(남재우, 김영준 2023). 이 방법은 단어의 공존 확률을 기반으로 각 문서의 토픽을 추론하고, 토픽 별 주요 단어를 찾아내는 원리를 가지고 있다. 예를 들어, 영어교육 관련 논문에서는 ‘English’와 ‘learn’이 함께 등장할 가능성이 크고, 통사론 관련 논문에서는 ‘phrase’와 ‘verb’가 함께 등장할 가능성이 크다. 토픽모델링에서는 이처럼 높은 빈도로 함께 등장하는 단어들이 서로 연관이 있다고 가정하며, 이들을 하나의 주제로 묶는다.

토픽 모델링에서 주로 사용되는 알고리즘은 LDA(Latent Dirichlet Allocation)로 문서가 여러 토픽으로 구성되어 있고, 각 토픽은 특정 단어들의 확률 분포로 나타난다는 가정을 기반으로 하고 있다. LDA 는 확률을 토대로 숨겨진 문서 집단의 의미론적 구조를 이해하는 방식으로, 문서-토픽 분포, 토픽-단어 분포를 추론하기 위해 알파(α), 베타(β)와 같은 하이퍼파라미터(hyperparameter)를 사용한다(Blei et al. 2010). 알파(α)값은 문서-토픽 분포를 결정하는 값으로 각 문서에 특정 토픽이 분포할 확률을 결정하며, 베타(β)값은 토픽-단어 분포를 결정하는 값으로 토픽에 단어가 분포할 확률을 결정한다.

2.4 선행연구

계량정보학 혹은 계량서지학 방법이 특정 학문 영역의 연구 동향 파악을 위해 사용되기 시작한 것은 2010 년대부터이다(Lei and Liu 2019a, Plonsky 2023). Plonsky(2023)에 따르면, 응용언어학 분야에서 계량서지학 연구는 2010~2014 년 4 편이었으나, 2015-2019 년에는 25 편, 2020~2023 년에는 40 편으로 지속적으로 증가하고 있다. Lei 와 Liu(2019a)는 언어학 분야의 계량서지학 분석의 시초가 되는 연구로 2005 년에서 2016 년 사이에 SSCI 에 등재된 응용언어학 관련 학술지 42 개의 총 10,028 개의 논문을 분석하여 주요 주제, 자주 인용되는 논문과 저자, 출판 국가를 추출하고, 해당 기간을 세 시기로 나누어 주요 주제의 증감을 파악하였다. 이후 (응용)언어학에서의 계량서지학적 연구는 System(Lei and Liu 2019b), Computer Assisted Language Learning(Goksu et al. 2022)과 같이 특정 학술지의 연구 동향을 파악하거나, 다중언어주의(Lin and Lei 2020), L2 문자 교정 피드백(Crosthwaite et al. 2022), ESP(Hyland and Jiang 2021), SLA(Zhang, 2020)와 같이 특정 연구 주제의 동향을 분석하는 방식으로 분화된다. 계량서지학적 분석은 특정 시기의 연구 주제와 관행을 체계적으로 파악할 수

있는 방법으로, 체계적 문헌분석, 메타 분석과 같이 지금까지 행해진 연구 결과를 분석하고 보고하는 통합적이며 성찰적인 연구로 볼 수 있다(Plonsky, 2023).

한국의 영어학 및 응용언어학 분야에서 수행된 계량적 방법으로 연구동향을 분석한 연구는 2000년대 후반부터 나타난다. 분석의 대상은 특정 학술지이거나 특정 연구 주제로 나누어진다. 이은주(2008)는 1998년 이후 영어교육과 응용언어학 분야 코퍼스와 관련된 국내외 94편의 논문을 대상으로 연구동향을 분석하였다. 그 결과 국내외 모든 연구가 어휘 관련 주제에 편중되어 있고, 정량적 분석을 주로 활용한 것으로 나타났다. 또한, 문어 코퍼스와 원어민 코퍼스가 주로 활용되었다. 이를 통해 정량적 연구와 정성적 연구의 균형, 적절한 구어 코퍼스의 조성 및 활용, 학습자 코퍼스 구축과 분석 등을 제안하였다. 차경환과 김영규(2010)는 1983~2008년까지 한국응용언어학회의 학술지인 《응용언어학》에 게재된 380편의 논문을 대상으로 연구동향을 분석하였다. 이를 통해 초기 학문적 정체성 수립 단계에는 이론언어학 분야의 연구가 수행되었고, 점차 멀티미디어나 코퍼스언어학 등을 통한 실용적 연구가 진행된 것을 발견하였다. 또한, 영어교육 분야의 연구가 주로 수행되어 오다가 한국어교육 분야 연구가 증가하는 추세를 발견하였다. 심영숙(2020)은 세계영어에 관한 158편의 논문을 대상으로 연구동향을 분석하였다. 그 결과 2000년대 중반 이후 세계영어 연구가 본격적으로 진행되어 점차 증가하고 있으며, 주로 영어교육 관련 학술지에 발표되고 있다는 사실을 발견했다. 또한 실증적이고 정량적인 연구가 이론적 연구보다 더 많이 수행되고 있는 것을 확인하였다. 이러한 분석결과를 바탕으로 특정 연구주제에 대한 치우침과 중복을 지적하며, 다양한 주제의 발굴이 필요하다고 제안하였다.

Netminer, 파이썬 등의 도구를 활용해 텍스트 마이닝을 수행한 계량서지학적 연구는 2020년대부터 본격적으로 나타나기 시작한다. 특히, LDA 토픽모델링과 시계열 분석의 방법이 자주 활용되고 있다. 신유선과 김양희(2020)는 2000~2019년 동안 《영어교육연구》 학술지에 게재된 814편의 논문 키워드를 대상으로 네트워크 분석을 수행하였다. 그 결과 20년간 주된 연구 주제는 의사소통능력, 동기, 읽기, 교사교육, 듣기였으며, 연구 주제의 확장과 주제 간의 연관성이 발견되었다. 권은영(2020)은 1990~2019년 동안 수행된 국내외 ESP 관련 400편의 논문 제목을 대상으로 텍스트 마이닝을 수행하였다. 단어의 빈도분석을 수행한 결과 국내외 공통으로 문자 커뮤니케이션, 교육과정, 어휘를 연구 주제로 선호하고 있었다. 하지만, 국내 ESP 연구에서는 코퍼스 분석을 가장 선호했고, 해외연구에서는 장르 및 요구분석과 사례 연구를 선호하고 있었다. 또한, 공학이 국내외 EAP 연구에서 가장 많이 다뤄진 학문 분야인 것으로 나타났다. 하지만, 신유선과 김양희(2020), 권은영(2020)의 연구는 분석 대상을 논문의 키워드나 제목과 같이 짧은 텍스트를 사용하였다는 한계가 있다.

박은희(2021a)는 2001~2020년 동안 수행된 교양영어 관련 논문 300여편의 초록을 대상으로 빈도분석, LDA 토픽모델링, 시계열분석을 수행하였다. 그 결과 대학 교양영어 관련 연구는 20년간 다양한 주제 영역에서 양적으로 크게 증가하였고, LDA 분석에서는 10개의 주제로 분류되었다. 시계열분석 결과를 통해 향후 학습자 요인과 관련된 연구가 증가할 것으로 예측하였다. 박은희(2021b)는 2001~2020년 간 수행된 영어쓰기 관련 논문 1,495편을 대상으로 LDA 분석과 시계열분석을 수행하였다. 분석결과 해당기간 영어 쓰기교육 관련 논문이 양적으로 크게 증가하였고, 토픽은 25개로 분류되었다. 시계열 분석결과 학습자 요인과 관련된 주제가 상승과 하강 주제 모두를 구성하고 있음을 발견하였다. 원용국과 김영우(2021)는 2000~2019년 동안 영어교육 주제의

논문 7,254 편의 초록을 대상으로 LDA 분석을 수행하였다. 상승 토픽과 하강 토픽을 파악하여 2000년대 이후 언어 형식 중심의 연구가 감소하고, 어휘를 중심으로 하는 데이터 기반 연구와 맥락을 중요시하는 정성적 연구가 늘어나고 있는 것을 발견하였다. 박신애(2022)는 1960~2021년까지의 영어 음운 관련 389편 논문의 제목을 대상으로 빈도분석, N-gram 분석, TF-IDF 분석, 매트릭스 분석, 네트워크 분석 등을 수행하였다. 분석결과 한국인 학습자와 한국어 음운 현상과 관련된 영어 음운 연구가 가장 많이 나타났고, 이어서 음운 인식, 학습자의 음운 인식 비교에 관한 연구가 많이 수행된 것을 발견하였다. 이를 통해 국내 음운 연구에서 억양 연구의 필요성을 제시하였다. 권민재(2024)는 1999~2023년 동안 수행된 독어학 관련 논문 1,968 편의 초록을 대상으로 언어 네트워크 분석을 수행하였다. 그 결과 응용 분야와 언어구조에 관한 연구가 큰 부분을 차지하지만, 의사소통행위와 대조분석 연구도 중요하게 나타나며, 연구분야가 세분화되고 인접 학문으로 확대되는 것을 확인하였다.

이러한 선행연구를 바탕으로 본 연구의 특징과 범위는 다음과 같다. 첫째, 특정 영어학 주제보다 《영어학》 학술지를 분석 대상으로 한다. 이는 《영어학》이 영어학 분야에서 학술지로 가진 중요성을 고려하여, 영어학 분야의 학문적 지형을 파악하기 위해서이다. 둘째, 지난 10년간 출판된 논문을 분석한다. 이는 최근의 연구 동향을 파악하기 위해서이다. 셋째, 상대적으로 긴 텍스트인 논문의 초록을 수집하여, 빈도분석, TF-IDF 분석, 토픽모델링을 실시한다. 빈도분석을 통해 주요 연구 주제를 가늠하고, 연도별로 TF-IDF가 높은 키워드를 추출하여 새롭고 독창적인 연구 주제를 찾아본다. 토픽모델링을 통해 주요 연구 주제와 영어학 하위 분야의 연구 동향을 파악한다.

3. 연구방법

3.1 연구설계

텍스트 마이닝은 일반적으로 데이터 선택, 데이터 정리, 데이터 변환(전처리), 데이터 마이닝, 결과해석의 5 단계로 수행된다(Fayyad et al. 1996). 본 연구에서는 이 절차에 따라 2014년부터 2023년까지 《영어학》에 게재된 논문 456 편의 초록을 대상으로 텍스트 데이터를 구축하여 분석에 적합한 데이터로 가공한 후 텍스트 마이닝을 수행하였다. 구체적으로 우선 10년간 빈도수가 높은 단어를 추출하고 이와 관련성이 높은 키워드를 추출하여 전반적인 연구 주제와 범위를 파악하고자 했다. 다음으로 연도별 중요 키워드 추출을 위해 TF-IDF 분석을 수행하였고, 주요 연구주제 및 하위 토픽 파악을 위해 LDA 분석을 수행하였다. TF-IDF 분석은 연도를 한 개의 집합단위로 구분해 수행하여 연구주제의 변화 추이를 살펴보고자 했다. LDA 분석은 전체 논문을 대상으로 수행하여 《영어학》의 연구주제를 파악하고자 했다.

TF-IDF는 각 단어의 중요성을 나타내는 지표를 제공하고, LDA는 문서의 숨겨진 토픽을 추출하여 문서 간의 구조를 이해하는 데 도움이 된다. TF-IDF 분석은 다른 데이터 마이닝 방법과 상호보완적으로 사용될 때 더 의미 있는 가치를 얻어낼 수 있다(Yang and Jia 2010). 본 연구에서는 두 가지 기법을 텍스트 마이닝의 주된 분석방법으로 사용하고, 파이(phi) 계수 산출을 통해 고빈도

출현 단어와 연관성이 있는 단어들을 살펴보았다. 분석은 R-4.3.3 프로그래밍 언어와 spacy, tidytext, tm, LDAvis 등의 패키지를 사용하여 수행하였다. 연구절차를 도식화하면 <그림 1>과 같다.



그림 1. 연구절차

3.2 데이터 구축 및 전처리

3.2.1 데이터 구축

한국학술지인용색인(KCI)에서 지난 10년간(2014~2023년) 《영어학》에 게재된 457 편의 논문을 검색하여 주요 서지사항을 내려 받고, 이 중 서평 1건을 제외한 456 편의 논문을 엑셀 파일로 정리하였다. 문헌 데이터는 저자명, 논문명, 키워드, 발행기관, 초록 등 다양한 필드로 구성되어 있지만, 텍스트 마이닝을 수행하기 위해서는 문장으로 구성된 초록 필드의 텍스트 데이터가 분석 대상으로 가장 적합했다. 따라서 각 문헌의 초록을 추출하여 코퍼스를 구축하고자 했고, 전반적인 검토를 통해 논문명과 초록의 내용이 일치하는지 확인하였다. 이 과정에서 180여 건의 초록에 서지정보(저자명, 논문명, 저널명 등)와 인용표기가 작성된 것을 발견하였고, 이는 불필요한 데이터이므로 삭제하였다. 이후 엑셀 데이터를 R(R Studio)로 가져와서 데이터 세트로 구축하고 초록 필드를 코퍼스로 구축하였다.

3.2.2 정제

데이터 정제(cleaning)는 데이터 분석에 불필요한 특수문자, 숫자, 불용어 등을 제거하여 텍스트 마이닝 결과의 정확성과 신뢰도를 높이기 위한 전처리 과정이다. 이 연구에서는 불필요한 공백과 숫자를 제거하였고, 관사, 전치사, 접속사, 대명사 등과 같이 자주 사용되지만, 큰 의미를 나타내지 않는 단어를 불용어로 선별해 제거하였다. 또한, 초록의 도입부에 상투적으로 등장하는 ‘This study’, ‘This paper’, ‘The purpose’ 등의 어구를 제거하여 정확한 분석을 위한 기반을 구축하였다.

3.2.3 정규화

정규화(normalization)는 표현 방법이 다른 단어를 통합하여 코퍼스 내의 단어를 일관된 형태로 변환하는 작업이다. 이 과정은 주로 어간 추출(stemming)과 표제어 추출(lemmatization) 두 가지 방법을 사용한다. 어간 추출은 단어의 변형된 형태에서 접사 등을 제거하여 어간을 추출하는 작업을 의미하며, 표제어 추출은 단어의 변형된 형태를 기본형으로 변환하는 작업이다. 이러한

정규화 작업은 코퍼스의 복잡성을 줄이고 자연어 처리의 효율성을 높이는 데 도움이 된다. 두 방법 모두 널리 사용되고 있지만, 어간 추출은 간혹 문맥 파악을 어렵게 할 수 있는 단점이 있다. 예를 들어, 'arguing'가 기본형 단어인 'argue'가 아닌 'argu'로 축소될 수 있다. 반면, 표제어 추출은 과거형, 복수형 등 다양한 단어 형태를 기본형으로 변환하므로 단어의 형태가 보존되며 의미 파악이 쉽다는 장점이 있다. 따라서 본 연구에서는 표제어 추출 방법을 사용하여 여러 형태의 단어를 기본형으로 통합했다.

3.2.4 토큰화

토큰화(tokenization)는 문장구조의 텍스트 데이터를 단어 단위로 분리하는 과정이다. 토큰화는 문장부호, 대소문자, 특수문자, 다양한 품사 등을 포함한 자연어 텍스트를 컴퓨터가 용이하게 인식하고, 계량적으로 정확하게 분석할 수 있도록 수행하는 전처리 방법이다. 이 연구에서는 456 편의 논문에 고유 ID 를 부여하고, 문장으로 구성된 초록을 한 개 단위의 단어로 분리하였다.

4. 분석결과

4.1 연구추세

《영어학》의 연구 추세는 <그림 2>와 같다. 2014-2020 년도까지 연간 40 건 미만의 논문이 게재되었는데, 2021 년부터 60 건 이상의 논문이 게재되고 있다. 이는 2020 년부터 수시접수와 발간 등 출판 정책을 변경하고 온라인 출판을 활성화하는 동시에, 2020 년 7 월 28 일에 SCOPUS 에 학술지가 등재되어 연구자의 선호도가 상승한 결과로 보여진다.

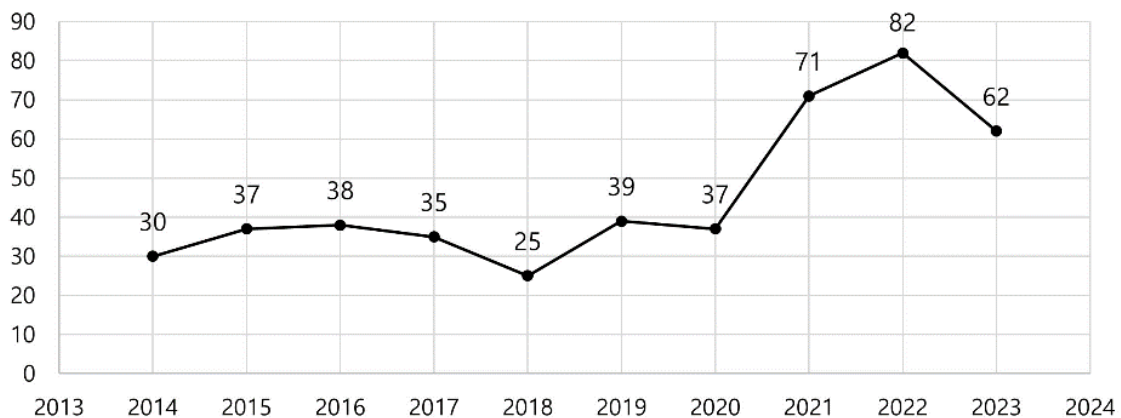


그림 2. 연구추세

4.2 빈도수 및 연관어 분석

지난 10 년간 《영어학》의 전반적인 연구 주제를 가늠하기 위해 고빈도 단어와 해당 단어와 상관관계가 높은 단어를 추출하였다. 빈도수가 높은 주요 단어 30 개와 밀접한 관계를 가진 단어를 알아보기 위해 단어 간의 파이계수를 산출하였고, 그 결과는 표 1 과 같다. 표 1 에서 30 개의 단어만 제시한 이유는 30 개의 단어를 통해 전반적인 연구주제의 동향을 파악하는데 무리가 없으며, 그 이상의 단어를 제시할 경우 정보전달의 효율성과 가독성을 저하시킨다고 판단하였기 때문이다.

파이계수는 변수 간의 상관관계를 측정하는 지표로 두 단어가 함께 사용될 확률을 수치로 표현한 것이다. 일반적으로 파이계수의 절대값이 0.4 보다 작으면 약한 상관관계, 0.4 에서 0.6 사이면 중간 정도의 상관관계, 그 이상이면 강한 상관관계라고 설명하지만(Schober et al. 2018), 0.3 을 중간 정도의 상관관계로 보는 기준도 있으며(Field 2009), 매우 강한 상관관계로 보는 견해도 있다(Akoglu 2018). 연구 분야나 목적에 따라 기준이 달라질 수 있으므로 분석 결과를 전체 맥락 속에서 해석하는 것이 중요하다.

표 1. 빈도수 상위 단어와 연관어

순위	단어	빈도	연관어
1	English	1,499	Korea(0.231), native(0.213), education(0.201), learn(0.195), elementary(0.169), teacher(0.15), curriculum(0.148), skill(0.139), school(0.136), interview(0.135)
2	learn	1,072	student(0.38), Korea(0.333), participate(0.316), level(0.307), proficiency(0.271), EFL(0.232), language(0.23), task(0.227), significant(0.218), experience(0.218)
3	student	669	learn(0.38), improve(0.346), participate(0.331), college(0.331), school(0.33), university(0.308), pedagogic(0.307), class(0.307), positive(0.294), significant(0.282)
4	Korea	564	learn(0.333), EFL(0.26), student(0.24), school(0.232), English(0.231), college(0.226), level(0.222), proficiency(0.211), native(0.188), write(0.175)
5	write	502	feedback(0.297), text(0.264), essay(0.253), argumentative(0.25), genre(0.244), student(0.235), overuse(0.226), academic(0.22), discipline(0.211), cohesion(0.208)
6	word	449	lexical(0.309), list(0.281), vocabulary(0.259), vowel(0.253), phonologic(0.241), token(0.224), consonant(0.223), stress(0.214), frequency(0.213), compound(0.206)
7	teacher	444	teach(0.443), elementary(0.413), service(0.395), school(0.366), class(0.359), interview(0.339), education(0.308), experience(0.304), quality(0.294), classroom(0.288)
8	level	381	proficiency(0.408), low(0.392), learn(0.307), student(0.26), significant(0.223), Korea(0.222), ability(0.199), develop(0.191), increase(0.191), total(0.187)
9	language	379	foreign(0.295), learn(0.23), global(0.205), text(0.192), benefit(0.184), classroom(0.17), context(0.165), ideology(0.157), education(0.151), AI(0.147)
10	test	349	taker(0.306), post(0.306), score(0.3), student(0.266), statistic(0.261), significant(0.26), ability(0.25), pre(0.246), divide(0.244), impact(0.243)
11	speak	331	native(0.548), ELF(0.232), speech(0.208), product(0.178), pitch(0.168), taker(0.165), actual(0.165), pronounce(0.165), produce(0.154), occur(0.154)
12	task	320	performance(0.23), learn(0.227), accurate(0.217), college(0.217), post(0.202), load(0.194), effect(0.19), test(0.189), assign(0.187), complete(0.181)
13	effect	302	significant(0.231), post(0.223), test(0.221), control(0.218), task(0.19), week(0.187), pre(0.185), learn(0.183), improve(0.177), prosody(0.169)
14	participate	272	student(0.331), learn(0.316), teacher(0.259), interview(0.258), pre(0.237), positive(0.23), program(0.217), university(0.208), active(0.208), experience(0.205)

15	read	262	comprehension(0.302), aloud(0.291), gloss(0.239), sentence(0.208), significant(0.194), measure(0.194), learn(0.192), grade(0.191), option(0.17), time(0.163)
16	significant	234	difference(0.387), statistic(0.308), student(0.282), test(0.26), low(0.259), measure(0.25), college(0.239), effect(0.231), score(0.225), level(0.223)
17	sentence	225	pronoun(0.319), cohmetrix(0.281), cohesion(0.277), neural(0.256), connect(0.25), syntax(0.246), human(0.244), readability(0.233), correct(0.223), count(0.216)
18	proficiency	220	level(0.408), low(0.369), learn(0.271), test(0.216), advance(0.215), Korea(0.211), increase(0.21), tendency(0.197), CEFR(0.197), predictor(0.189)
19	vocabulary	217	pedagogic(0.282), gloss(0.274), load(0.272), student(0.271), word(0.259), gain(0.25), list(0.25), family(0.248), implicate(0.246), productive(0.242)
20	school	211	middle(0.487), teacher(0.366), student(0.33), elementary(0.3), curriculum(0.293), secondary(0.248), Korea(0.232), education(0.207), teach(0.207), depth(0.204)
21	verb	209	noun(0.338), morpheme(0.334), count(0.287), argument(0.285), type(0.276), syntax(0.261), cohmetrix(0.26), readability(0.26), pronoun(0.257), phrase(0.243)
22	type	205	verb(0.276), gloss(0.182), script(0.182), difficult(0.167), adjective(0.165), predicate(0.164), CP(0.159), telicity(0.158), yield(0.158), cohmetrix(0.154)
23	corpus	204	frequency(0.36), connect(0.319), token(0.319), lexical(0.27), coverage(0.262), cohesion(0.241), America(0.234), marker(0.225), frequent(0.217), BNC(0.213)
24	process	202	psycholinguistic(0.225), reaction(0.206), BERT(0.192), human(0.187), sentence(0.18), neural(0.179), close(0.175), seek(0.175), revision(0.168), semantically(0.168)
25	develop	196	quality(0.365), elementary(0.268), positive(0.246), teach(0.236), interview(0.228), teacher(0.228), program(0.224), content(0.219), professional(0.21), student(0.209)
26	data	193	collect(0.325), collection(0.261), note(0.243), quality(0.222), interview(0.216), corpus(0.2), field(0.19), journal(0.174), data(0.168), lesson(0.166)
27	teach	192	teacher(0.443), quality(0.368), elementary(0.353), lesson(0.316), service(0.31), classroom(0.295), education(0.283), implicate(0.254), prepare(0.236), professional(0.236)
28	syntax	190	semantic(0.335), cohmetrix(0.298), count(0.274), verb(0.261), move(0.249), readability(0.248), sentence(0.246), wh(0.239), argue(0.233), clause(0.23)
29	education	188	curriculum(0.325), professional(0.313), teacher(0.308), elementary(0.308), quality(0.3), teach(0.283), service(0.279), program(0.257), content(0.256), experience(0.23)
30	difference	185	significant(0.387), statistic(0.216), score(0.199), discuss(0.182), positive(0.181), measure(0.167), student(0.164), divide(0.16), compare(0.159), detailed(0.154)

빈도수와 연관되는 어휘를 바탕으로 《영어학》에 지난 10년간 출판된 연구의 경향을 정리하면 다음과 같다.

첫째, 영어교육 및 제2언어 습득 및 발달과 관련된 연구가 주도적이다. ‘learn’, ‘student’, ‘teacher’, ‘school’, ‘education’ 등과 같은 교육과 관련된 단어의 출현 빈도가 높으며, 가장 빈도수가 높은 ‘English’, ‘develop’의 연관어도 교육과 관련된 것이 많다.

둘째, 언어 기능(skill)과 관련된 ‘write’, ‘speak’, ‘read’ 단어도 빈도수가 높아 영어교육 연구의 우세함을 암시하지만, 연관어를 보면 영어학 내 다른 분야와의 연관성을 유추할 수 있다. 쓰기의 경우, 영어교육과 관련된 ‘feedback’은 물론 ‘text’, ‘genre’, ‘cohesion’과 상관관계가 높아 담화 연구와의 관련성을 유추할 수 있다. 말하기는 ‘native’, ‘ELF’처럼 공용어로서의 영어라는 담화 혹은 사회언어학 주제와의 연관성은 물론, ‘pitch’, ‘pronounce’와 같은 단어에서 음성음운론과의 접합점을 찾을 수 있다. 다만, 읽기와 듣기같은 이해 영역의 경우, 쓰기와 말하기 같은 표현 영역과 다른 양상을 보인다. 읽기의 경우 빈도수가 상대적으로 낮으며, 연관어도 ‘comprehend’, ‘measure’과 같이 영어교육 연구와 관련성이 높다. 듣기는 상위 30개 고빈도 단어 목록에 포함되지 않았다.

셋째, 언어 형식(form)으로는 ‘word’, ‘sentence’, ‘vocabulary’, ‘verb’이 포함되었으며, 영어학 분야로는 ‘syntax’가 고빈도 어휘로 추출되었다. 일반언어학 분야 중 통사론에 대한 연구가 많이 진행되었음을 유추할 수 있으나, ‘word’의 연관 어휘에 ‘vowel’, ‘phonologic’, ‘consonant’, ‘stress’가 포함된 것에서 보듯이, 음성음운론 분야도 활발히 연구되고 있는 것으로 보여진다.

넷째, 영어학의 하위분야를 명시적으로 나타내는 또다른 어휘는 ‘corpus’와 ‘process’이다. 두 단어는 각각 코퍼스언어학과 심리언어학을 지표한다. 특히, 코퍼스언어학에 대한 관심은 고빈도 어휘로 포함된 ‘word’, ‘vocabulary’의 연관어를 통해서도 살펴볼 수 있다.

다섯째, 연구자료 및 방법과 관련된 어휘로는 ‘test’, ‘task’, ‘effect’, ‘participate’, ‘significant’, ‘data’, ‘difference’ 등이 있다. 해당 단어와 연관성이 높은 어휘를 보았을 때 유추할 수 있는 연구 설계는 비교집단과 통제집단을 설정하는 ‘실험 혹은 준실험 연구(experimental and quasi-experimental research)’ 혹은 사전-사후 시험 결과를 통계적으로 비교하는 ‘처치 효과 측정’(measuring the effect of treatment) 연구이다(Mackey and Gass 2022).

4.3 TF-IDF 분석

전체 논문을 연도별로 그룹화하고, TF-IDF 수치에 근거하여 연도별 중요단어를 추출하였다. TF-IDF 수치는 단어의 빈도수와 해당 단어가 나타나는 논문의 수를 기반으로 계산되므로, 특정 연도에서 중요하게 여겨지는 단어를 통계적으로 식별할 수 있다. 이를 통해 《영어학》에서 일반적으로 다루어지는 주제 외에 상대적으로 특징적인 연구를 파악할 수 있다. 예를 들면, 표 2 의 2021 년 단어들은 다른 연도와 비교해 2021 년에 출판된 논문에서 높은 빈도로 나타나므로, 2021 년을 특징짓는 키워드로 간주할 수 있다. 가령 ‘chatbot’ 단어를 통해 당시 인공지능 기술의 급격한 발전이나, 이와 관련된 새로운 연구 주제의 등장을 유추할 수 있다. 따라서, TF-IDF 수치가 큰 단어들의 연도별 추세를 통해 학술지의 연구 동향과 관심사의 변화를 추론할 수 있다. 다만, TF-IDF 분석을 통해 추출된 단어들은 통계적 특징을 나타내는 것일 뿐이며, 연구의 질을 나타내지는 않는다. 또한, 2014~2020 년까지의 기간 동안 연도별 40 개 미만의 논문이 게재되었고, 이러한 제한으로 인해 한 개의 논문에서 다빈도로 나타난 단어의 TF-IDF 값이 높게 나타난 경향이 있을 수 있다.

연도별 TF-IDF 수치가 높은 키워드와 관련된 논문을 통해, 해당 연구의 독창성을 가늠해보면 다음과 같다. 다만, TF-IDF 값 기준 상위 키워드 중 축약어가 추출된 경우(예를 들어, ‘TC’, ‘FonF’, ‘VR’ 등), 해당 연구 주제의 독창성에 기인할 수도 있지만, 축약어의 특이성을 반영할 수도 있음에 유의할 필요가 있다.

표 2. 연도별 TF-IDF 값 상위 10 개 단어

연도	주요 단어	TF-IDF	연도	주요 단어	TF-IDF	연도	주요 단어	TF-IDF
2014	annotation	0.600	2018	tie	0.702	2022	gloss	0.731
	glottalize	0.500		tonic	0.702		space	0.501
	telephone	0.500		talent	0.585		KS	0.439
	coercion	0.400		anaphora	0.550		autonomy	0.321
	DC	0.400		checker	0.490		CT	0.292
	OV	0.400		EPP	0.490		hiatus	0.292
	blend	0.333		flip	0.489		COVID	0.251
	option	0.318		quantifier	0.468		MCG	0.251
	enlightenment	0.300		neoliberal	0.409		war	0.251
	interviewer	0.300		creative	0.372		metaphor	0.233
2015	comma	0.487	2019	email	0.612	2023	frame	0.424
	summary	0.442		tutor	0.498		intervenor	0.417
	dead	0.423		VR	0.476		suicide	0.417
	imperfective	0.423		cohmetrix	0.356		aptitude	0.341
	perfective	0.423		conference	0.340		multiword	0.341
	rhythm	0.423		liquid	0.333		particle	0.278
	sibilant	0.423		collocation	0.285		metaphor	0.272
	DM	0.355		freewrite	0.272		ChatGPT	0.266
	editorial	0.355		Shakespeare	0.272		translingualism	0.266
	palatalization	0.355		educator	0.271		generic	0.265
2016	TC	0.619	2020	FonF	0.619			
	collocation	0.485		tutor	0.567			
	ASC	0.387		film	0.387			
	encoding	0.309		PolP	0.387			
	familiar	0.309		preprocess	0.387			
	LSI	0.309		TCU	0.387			
	pun	0.309		backchannel	0.352			
	wh	0.287		absolute	0.339			
	rubric	0.283		CEFR	0.283			
	implicit	0.277		complement	0.275			
2017	determine	1.358	2021	chatbot	0.716			
	notion	0.621		song	0.409			
	epenthesis	0.534		NS	0.358			
	Acc	0.509		authorial	0.256			
	passivization	0.509		dictogloss	0.256			
	EF	0.424		tagalog	0.256			
	numeral	0.415		generic	0.250			
	sort	0.415		mindset	0.250			
	force	0.355		AI	0.227			
	CDMs	0.339		engage	0.221			

2014년에는 ‘annotation’(0.600), ‘glottalize’(0.500), ‘telephone’(0.500) 등의 단어가 0.5 이상의 TF-IDF 값을 나타내고 있다. ‘annotation’은 비즈니스 글쓰기에서 학생의 주석에 대한 교사의 피드백이 글쓰기 향상에 미치는 영향에 대한 연구(Lee 2014), ‘glottalize’은 벅아이 코퍼스를 이용하여 형태적, 사회문화적, 음성음운적 요인에 따라 치경폐쇄음과 어말 치경비음이 연속해서 나타나는 양상을 분석한 연구(신승훈 2014), ‘telephone’은 전화 영어회화에 나타난 대화의 상호작용 양상을 살펴본 연구(조선, 신동일 2014)에서 추출되었다. 2015년에 추출된 키워드의 TF-IDF 수치는 다른 연도에 비해 상대적으로 낮은데, ‘comma’(0.487)는 쉼표의 4 가지 유형에 대한 한국 대학생의 인식과 이가 글쓰기 과업에 미치는 영향을 살펴본 연구(Kwon and Park 2015), ‘summary’(0.442)는 텍스트 유형과 길이에 따른 과업의 인지적 복잡성이 요약 글쓰기에 미치는 영향을 살펴본 연구(Park 2015)와 관련된다. 2016년에는 ‘TC’가 TF-IDF 값이 0.619로 높았다. ‘TC’는 Tough-Construction의 축약어로 touch 구문을 통사적으로 분석한 연구(Kim 2016)에서 추출되었다. ‘collocation’(0.485)도 상대적으로 높았는데, 한국 영어학습자를 위한 언어 표준화 시험 개발 및 타당화에 관한 연구(Han and Lee 2016)와 관련된다. 2017년에는 ‘determine’(1.385), ‘notion’(0.621) 단어가 높은 TF-IDF 값을 나타내고 있다. ‘determine’과 ‘notion’은 학술논문에서 일반적인 의미로 사용될 수 있는 단어이다. 해당 단어는 2017년 게재된 35 개의 논문 중 각각 4 개 논문의 초록에서 나타났지만, ‘determine’의 TF-IDF 값이 매우 높게 나타난 주된 이유는 Yang(2017)의 연구 때문이다. Yang(2017)은 촘스키 최소주의이론의 결정론적 문제를 derivational null hypothesis라는 개념(notion)으로 해결하려는 이론적 시도이다. Yang(2017)의 논문 초록에는 ‘determinism’과 ‘notion’이 반복적으로 동시에 출현하며, TF-IDF 값을 고려했을 때 상당히 독창적인 연구로 보여진다. 한편, 0.534의 TF-IDF 값을 나타내는 ‘epenthesis’는 미국 영어의 [u]-약화와 [w]-삽입 현상을 분석한 연구(황보영식 2017)에서 추출되었다. 2018년에는 ‘tie’(0.702), ‘tonic’(0.702), ‘talent’(0.585) 등의 단어가 높은 TF-IDF 값을 나타내고 있다. ‘tie’는 L2 읽기에서 대용어(anaphora) 관련 전략이 영어능속도와 학문적 배경, 대용어 유형, 대용어 결속의 추론적 복잡성에 따라 어떠한 차이가 나는 지를 분석한 연구(Choi et al. 2018)에서, ‘tonic’은 한국인 영어 화자를 발화이해도와 영어 악센트에 따라 3 개의 집단으로 나눠 영어 발화에서의 휴지(pause)와 제 1 강세(tonic stress)의 사용 양상을 분석한 Yang(2018)의 연구에 기인한다. ‘tie’와 ‘tonic’은 해당 연도의 다른 단어들과 비교해 TF-IDF 값이 크므로, 해당 논문은 독창성을 갖고 있거나 많이 연구되지 않는 주제일 가능성이 높다. ‘talent’는 신문기사를 통해 글로벌 인재(global talent)의 영어능력에 관한 담론을 분석한 연구(신동일 2018)와 관련이 있다. 2019년에는 ‘email’(0.612), ‘tutor’(0.498), ‘VR(Virtual Reality)’(0.476)의 단어가 높은 TF-IDF 값을 나타내고 있다. ‘email’은 실제 비즈니스 상황에서 영어 원어민과 비원어민이 작성하는 요청 이메일 작성 양상을 비교한 연구(박상희, 전지현 2019)에서, ‘tutor’은 글쓰기 튜터의 문자 피드백이 한국인 영어 학습자의 글쓰기 정확성, 복잡성, 유창성에 미치는 영향에 관한 연구(Cho 2019)에서, ‘VR’은 VR을 활용한 영어 듣기 평가의 타당성을 검증한 연구(Lee 2019)에서 추출되었다. 2020년에는 ‘FonF(Focus on form)’(0.619), ‘tutor’(0.567)의 단어가 높은 TF-IDF 값을 나타낸다. ‘FonF’는 블렌디드 학습 상황에서 형태초점교수방법의 효과성을 살펴본 연구(Kang and Lee 2020)에서, ‘tutor’는 초보 글쓰기 튜터의 학생과의 상호작용이 한 학기 동안 어떻게 변화하는 지를 숙련된 글쓰기 튜터와 비교 분석한 연구(Cho et al. 2020)에서 추출되었다.

2021년부터는 《영어학》이 온라인 출판으로 전환되고, SCOPUS에 등재되면서 발행되는 논문수가 증가하기 시작하였다. 2021년에는 ‘chatbot’(0.716)이 높은 TF-IDF 값을 나타내고 있다.

‘chatbot’은 인공지능 기반 챗봇이 제 2 언어 대화상대로 적합한지를 챗봇과의 실제 대화 자료를 통해 분석한 연구(Sin et al. 2021), AI 를 활용한 영어 교수학습 연구의 언어 네트워크를 분석한 연구(권은영 2021), 요구분석을 바탕으로 항공승무원을 위한 챗봇을 개발하고 시범 적용한 결과를 분석한 연구(Yang and Kim 2021), 챗봇이 EFL 학습자의 말하기 능력과 정의적 영역에 미치는 영향을 분석한 연구(Kim et al. 2021) 등 4 개의 논문 초록에서 출현한다. ‘chatbot’은 해당 연도에서는 빈도수가 높지만, TF-IDF 의 원리를 고려했을 때, 전체 문서 집합에서는 흔하게 등장하지 않는 단어이므로, 2021 년 출판된 논문에서 중요하게 출현한 주제로 볼 수 있다. 다만, 동일한 저자가 챗봇과 관련된 주제로 해당 연도에 복수의 논문을 출간한 것도 영향을 미쳤을 수 있다. 2022 년에는 ‘gloss’(0.731)과 ‘space’(0.501)이 높은 TF-IDF 값을 나타내고 있다. ‘gloss’의 경우, 다양한 유형의 주석이 L2 학술 어휘 습득에 미치는 영향에 관한 세 편의 연구(Kang 2022a, 2022b, 2022c)에 기인한다. 이 주제도 동일한 저자가 세편의 논문을 한해에 출간한 것이 TF-IDF 값을 높이는 데 기여했을 수 있다. ‘space’는 학습 어휘 목록을 한번에 제시하는 것과 일정한 개수로 나누어 제시하는 것(1 개씩, 5 개씩, 10 개씩), 즉, 제시 어휘 목록의 간격(spacing)이 단어 및 관용어 학습에 미치는 영향을 파악한 연구 (Choi and Yun 2022)에서 추출되었다. 2023 년 논문은 2015 년과 같이 TF-IDF 값이 0.5 이상인 키워드가 추출되지 않았지만, ‘frame’(0.424), ‘intervenor’(0.417), ‘suicide’(0.417)가 상대적으로 높은 TF-IDF 값을 나타냈다. ‘frame’은 L2 학습자 코퍼스를 이용하여 학술 글쓰기에서 어휘 프레임(lexical frame) 사용과 그 속에서 정관사 사용 양상이 글쓰기 수준에 따라 어떠한 차이가 나는 지 분석한 연구(Jin 2023)에서, ‘intervenor’는 영어 it 분열구문(English it-cleft)에서 filler 와 gap 사이에 들어간 intervenor 의 지시 형태(referential form)가 문장 처리에 미치는 영향을 분석한 연구(Yu 2023)에서, ‘suicide’는 감성분석을 사용하여 자살 유서를 감지하는 알고리즘을 제안한 연구(Lee 2023)에서 추출되었다.

TF-IDF 분석 결과, 연도별로 주요 단어가 거의 다르게 나타나며 영어학 분야의 연구 주제가 매우 다양함을 알 수 있다. 높은 TF-IDF 값을 나타내는 키워드를 바탕으로 해당 논문 주제를 살펴보았을 때도 영어학의 다양한 영역에서 새로운 주제가 탐색되고 있음을 확인할 수 있다. 이러한 다양한 주제 속에서 주목할 만한 부분은 인공지능을 비롯한 과학기술의 발전에 따른 키워드가 2020 년대를 즈음하여 나타나기 시작한 것이다. ‘VR’이 2019 년 상위 키워드에 포함되었으며, 2021 년에는 ‘chatbot’, ‘AI’, 2023 년에 ‘ChatGPT’가 포함되었다. 이를 통해, 영어학 분야에서도 최근 과학기술의 발전에 따른 새로운 연구 주제 및 방법이 출현하고 있음을 알 수 있다. 2022년 ‘COVID’(0.251)가 중요한 키워드에 포함된 것도 영어학 분야가 사회 현상을 반영하고 있음을 시사한다. 한편, 전반적인 연구 경향을 보여주는 빈도수와 연관어 분석에서 나타나지 않았던 사회언어학 분야와 관련된 키워드가 TF-IDF 값 상위 키워드에 포함된 것도 주목할 만하다. 그 값이 상대적으로 높지는 않았지만 2018 년 ‘neoliberal’(0.409)과 2023 년 ‘translingualism’(0.266)이 주요 키워드로 나타난 것은 2010 년대 해외 사회언어학과 응용언어학 분야에서 활발히 논의되던 정치경제학 관점(Park and Lo 2012)과 횡단언어하기 이론(Li 2018)의 영향인 듯하다.

4.4 LDA 기반 토픽 모델링 결과

4.4.1 자료 변환과 토픽 수 산출

LDA 토픽 모델링을 수행하기 위해서는 전처리와 토큰화가 완료된 코퍼스를 DTM(Document-Term Matrix) 형태의 자료로 변환해야 한다. DTM 은 각 문서와 단어의 출현 빈도를 표현한 행렬로 서로 다른 토픽 수를 반복적으로 대입하며 LDA 알고리즘이 적절한 토픽 수를 추출하기 위한 입력 데이터로 사용된다. 산출된 값에 대한 적합도 지수 기준은 Deveaud et al.(2014), Griffiths and Steyvers(2004), Cao et al.(2009), Arun et al.(2010)이 제시한 방법들이 있다. Griffiths and Steyvers(2004)와 Deveaud et al.(2014)는 토픽 수의 최대값을 기준으로 선택하는 방법을 사용하고, Cao et al.(2009)와 Arun et al.(2010)은 토픽 수의 최소값을 기준으로 선택하는 방법을 사용한다. 본 연구는 Griffiths and Steyvers(2004)의 최대값과 Cao et al.(2009) 및 Arun et al.(2010)의 최소값이 교차하는 지점을 기준으로 최적의 토픽 수를 40 개로 결정하였다.

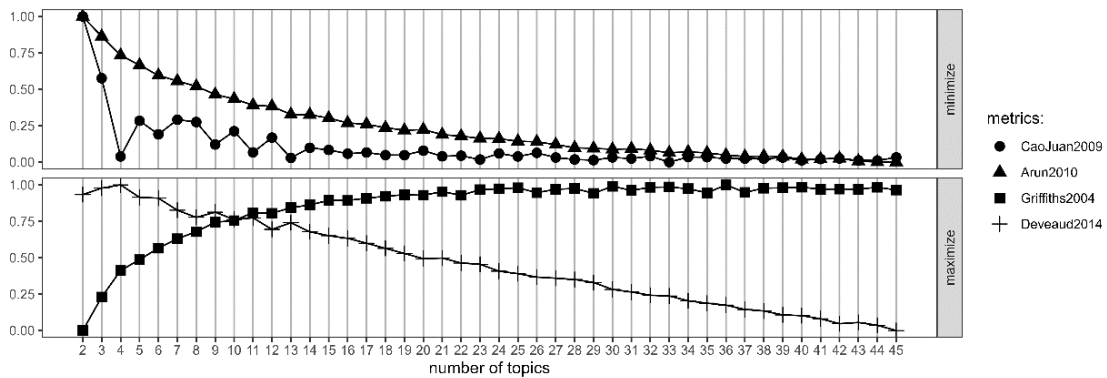


그림 3. 최적 토픽 수 산출

이후 pyLDAvis 패키지를 사용하여 LDA 분석결과를 시각화 하였다. pyLDAvis 는 다차원 척도법(multidimensional scaling)을 통해 토픽 간의 거리를 시각적으로 표현하여, 토픽 간의 관계 구조를 쉽게 파악할 수 있게 해준다. 각 토픽을 나타내는 원은 서로 가까울수록 관련성이 높으며, 원의 면적은 전체 문서에서 해당 토픽의 비중을 나타낸다. 이를 토대로 <그림 4>를 살펴보면 1~40 까지의 각 원은 비슷한 거리를 두고 분포하고 있으며, 서로 겹치거나 군집되어 있지 않음을 확인할 수 있다. 이는 각 토픽이 독립적으로 구성되었음을 시사한다. 또한, 원의 크기는 약간의 차이가 있지만, 두드러지게 큰 원이 없으므로 어느 한 토픽에 비중이 치우쳐 있지 않음을 알 수 있다. 종합하면 《영어학》에 게재된 456 편의 논문은 서로 독립된 40 개의 소주제로 구성되어, 다양한 주제 분야가 연구되고 있는 것으로 평가할 수 있다.

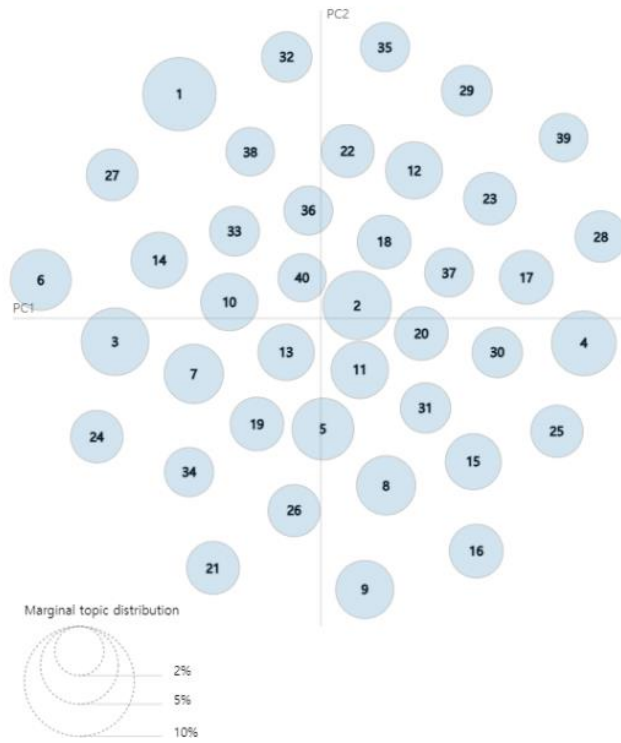


그림 4. LDA 분석 결과의 시각화

4.4.2 토픽별 주제명 추출

2014년부터 2023년까지 《영어학》에 게재된 456편의 논문을 대상으로 LDA를 수행한 결과는 표 3과 같다. 각 토픽별로 10개의 주요 단어를 추출하였고, 단어의 의미를 반영해 토픽을 대표하는 주제명을 작성하였다.

표 3. 토픽 주제와 주요 키워드

토픽번호	주제명	주요 키워드
TP1	읽기 전략 및 측정	read, strategy, comprehension, skill, question, input, measure, ability, vary, student
TP2	공용어로서의 영어(EFL) 의사소통	communicate, ELF (English as a Lingua Franca), understand, speak, conversation, interact, business, international, strategy, repair
TP3	언어 처리	process, model, language, time, human, sentence, linguistic, expect, compare, neural
TP4	형태초점 (Focus on Form)	form, focus, noun, adjective, mismatch, explore, phrase, voice, relation, context
TP5	낮은 영어 능숙도 학습자	level, proficiency, low, EFL, test, relate, collocation, increase, score, anxiety
TP6	초·중등 영어교육	English, school, education, Korea, middle, select, primary, secondary, quality, grader
TP7	통사구조 및 문법 전이	syntax, structure, evidence, grammar, condition, examine, reflexs, similar, transfer, role

TP8	리터러시 실천 및 담론	language, linguistic, book, context, practice, discourse, picture, critical, ideology, child
TP9	VP 생략 현상	clause, VP, ellipsis, wh, move, account, phrase, argue, follow, property
TP10	과업 수행	task, effect, complex, performance, accurate, term, participate, measure, plan, affect
TP11	부정문 및 해석	sentence, negative, subject, adverb, interpret, answer, context, event, position, judge
TP12	대학 영어 수업	student, class, improve, effect, post, university, EFL, design, score, significant
TP13	영어 듣기 인식 및 어려움	listen, perceive, active, Korea, difficult, participate, cues, adult, CSAT(College Scholastic Ability Test), MABL(Mobile-Assisted Blended learn)
TP14	연구 방법 및 동향	method, focus, topic, conduct, paper, frequent, trend, quantity, behavior, country
TP15	학술 글쓰기	write, academic, quality, essay, compare, instruct, implicate, genre, frame, specific
TP16	논항 구조	type, predicate, verbal, complete, root, accomplish, role, manner, action, derive
TP17	영어학습 동기	learn, motive, participate, factor, experience, EFL, college, week, autonomy, component
TP18	담화 기능 및 장치	function, category, pragmatic, marker, semantic, information, discourse, modal, dynamic, utterance
TP19	비원어민 화자로 영어 말하기	English, Korea, speak, native, product, learn, interpret, compare, influence, sensitive
TP20	자료 수집 및 구성	data, engage, relationship, collect, utilize, comment, platform, content, relate, comparative
TP21	초분절 자질	speech, stress, word, pattern, blend, rate, pronounce, vary, prosody, pause
TP22	음운론적 변이	vary, produce, phonologic, cluster, pattern, apply, rule, frequent, KS, delete
TP23	영어교육과정	language, curriculum, aspect, culture, revise, implement, context, competency, foreign, interculture
TP24	영주제(null topic)	feature, effect, derive, move, theory, null, topic, determine, object, cleft
TP25	영어 교사 교육	teacher, English, teach, develop, participate, program, elementary, classroom, education, service
TP26	언어 능숙도 차이	significant, difference, pair, term, relate, statistic, low, college, performance, advance
TP27	수험자 정체성 및 경험	test, taker, identity, prepare, global, narrative, include, corpus, language, focus
TP28	교과서 어휘 분석	word, frequency, corpus, textbook, lexical, list, include, elementary, unit, token
TP29	구문 문법	verb, construe, semantic, complement, idiom, construction, pattern, strong, argument, particle
TP30	영어 온라인 튜터링	tutor, reveal, interact, online, examine, context, presence, tutee, regulation, aim
TP31	음성학적 차이	vowel, duration, stop, word, acoustic, contrast, æ, follow, voiceless, phonetic
TP32	(영미)영어 형태	English, acquire, America, preference, British, compound, internal, examine, suffix, relate
TP33	코퍼스 분석	data, achieve, deep, corpus, note, direct, involve, sentiment, progress, suicide
TP4	오류 평가 및 피드백	assess, error, feedback, grammar, question, explore, provide, orient, address, correct

TP35	어휘 교수/학습	learn, vocabulary, instruct, knowledge, effect, test, explicit, expose, gain, involve
TP36	담화 내 지시어	complex, pronoun, sentence, corpus, lexical, reference, NP, cohesion, connect, measure
TP37	은유를 활용한 담화분석	metaphor, source, domain, discourse, reveal, term, usage, specific, COVID, affix
TP38	AI 활용 영어교육	translate, text, AI, output, English, chatbot, purpose, target, google, ESP(English for specific purpose)
TP39	분석 항목 및 정보 선택	item, information, choice, journal, abstract, diagnostic, addition, conduct, space, section
TP40	섬(island) 효과	approach, argue, claim, island, property, provide, propose, generic, define, review

명명한 토픽의 주제명을 <<영어학>> 편집위원회 구성⁴에 따라 음성음운론, 통사론, 의미·담화, 응용언어학·영어교육 분야로 범주화하고, 해당 토픽에 속하는 논문 수를 산출한 결과는 표 4 와 같다.⁵

토픽별 논문 수 산출은 LDA 분석에서 산출되는 감마값(gamma)을 활용했다. LDA 분석에서는 각 논문이 특정 토픽에 속할 확률(gamma)을 계산하므로, 계산 결과에 따라 할당된 논문을 집계하는 방법이다. 한 편의 논문이 여러 토픽에 속할 확률이므로, 논문이 여러 토픽에 중복해서 속할 수 있다.

표 4. 영어학 하위 영역에 따른 분류

범주	토픽번호	주제명	해당 논문 수
응용언어학·영어교육	TP1	읽기 전략 및 측정	12
	TP4	형태초점	11
	TP5	낮은 영어능숙도 학습자	14
	TP6	초·중등 영어교육	9
	TP12	대학 영어 수업	28
	TP13	영어 듣기 인식 및 어려움	10
	TP15	학술 글쓰기	15
	TP17	영어학습 동기	18

⁴ <http://www.kasell.or.kr/18>

⁵ 산출된 토픽 중 TP10(과업수행), TP14(연구 방법 및 동향), TP20(자료 수집 및 구성), TP32((의미)영어 형태), TP39(분석 항목 및 정보 선택)은 토픽 범주화에서 제외되었다. TP20 과 TP39 는 본 연구의 자료인 초록이 연구 내용 뿐만 아니라 연구 방법까지 포함하는 장르적 특징 때문에 도출된 것으로 보여진다. TP10 은 과업의 의미가 모호하여 제외하였다. 과업은 영어교육에서 사용하는 의사소통과업을 의미할 수도 있고, 실증연구에서 실험에 사용된 과업을 의미할 수도 있다. TP14 는 영어학 하위분야이기보다 연구 유형에 속한다. 본 연구의 초록을 토픽 모델링하면 TP14 에 포함될 가능성이 높다. TP32 은 영어학의 유의미한 하위분야이나 분류에 사용된 범주에 포함되지 않아 제외하였다. 분류가 모호한 토픽을 제외했음에도 불구하고, 여전히 특정 토픽이 해당 범주에 포함된 것에 이견이 있을 수 있다. 예를 들어, TP4(형태초점)은 영어교육에서 주로 논의되는 Focus on Form 을 의미할 수도 있지만, 언어 형식에 초점을 두고 진행했다는 연구 방법에 대한 의미일 수도 있다. TP33(코퍼스분석)의 경우도 문법, 어휘, 담화 등의 다양한 언어 형식을 분석 대상으로 삼았을 가능성이 있다. TP3(언어처리)는 심리언어학 분야와 밀접한 관련이 있으며, TP29(구문문법)도 전통적인 의미의 통사 이론과 차이가 있다. TP2(공용어로서의 영어(ELF) 의사소통)은 의사소통 측면을 고려하여 담화 연구로 보았지만, 영어교육 혹은 사회언어학적 함의가 크다. 이러한 모호함에도 불구하고 해당 토픽을 범주에 포함시킨 것은 엄밀한 학문적 범위보다 <<영어학>>의 편집위원회 구성을 바탕으로 통상적인 의미의 영어학 연구 내용과 범위를 고려했기 때문이다. 이후 논의하듯이, 이는 계량적으로 산출된 결과를 바탕으로 주제명명 및 범주화해야 하는 계량서지학 방법의 한계이기도 하다.

	TP19	비원어민 화자로 영어 말하기	17
	TP23	영어교육과정	11
	TP25	영어 교사 교육	25
	TP26	언어 능숙도 차이	6
	TP27	수험자 정체성 및 경험	10
	TP28	교과서 어휘 분석	14
	TP30	영어 온라인 튜터링	4
	TP33	코퍼스 분석	5
	TP34	오류 평가 및 피드백	10
	TP35	어휘 교수/학습	20
	TP38	AI 활용 영어교육	13
	토픽 수	19 (47.5%)	252 (55.3%)
통사론	TP3	언어 처리	11
	TP7	통사구조 및 문법 전이	14
	TP9	VP 생략 현상	39
	TP11	부정문 및 해석	9
	TP16	논항 구조	4
	TP24	영주제(null topic)	7
	TP29	구문 문법	14
	TP40	섬(island) 효과	10
	토픽 수	8 (20%)	108 (23.7%)
의미·담화	TP2	공용어로서의 영어(ELF) 의사소통	10
	TP8	리터러시 실천 및 담론	11
	TP18	담화 기능 및 장치	10
	TP36	담화 내 지시어	11
	TP37	은유를 활용한 담화분석	6
	토픽 수	5 (12.5%)	48 (10.5%)
음성음운론	TP21	초분절 자질	10
	TP22	음운론적 변이	9
	TP31	음성학적 차이	16
	토픽 수	3 (7.5%)	35 (7.7%)

토픽모델링 결과를 바탕으로 지난 10 년간 《영어학》의 연구 주제 동향을 다음과 같이 유추할 수 있다. 우선, 토픽의 개수와 주제명을 보았을 때, 다양한 영어학 영역에 대한 연구가 진행되고 있음을 발견할 수 있다. 하지만 연구 분야에 따라 분류해보았을 때, 응용언어학·영어교육 분야가 주제와 논문 수에서 가장 우세하였다. 일반영어학 분야에서는 통사론 연구가 상대적으로 다양한 주제로 많이 수행되었다. 또한, 토픽에 할당된 논문 수를 바탕으로 각 영역 내 주요 연구 주제를 보았을 때, 응용언어학·영어교육 분야에서는 대학영어(TP12), 영어교사교육(TP25), 어휘교수학습(TP35)이, 통사론 분야는 VP 생략 현상(TP9)에 대한 연구가 많았다. 통사론 분야에 포함된 주제는 심리언어학과 연관성이 강해 보이거나(TP3), 전통적인 의미의 통사 이론과 다른 주제(TP29)도 포함되어, 학문의 경계가 확장하고 있음을 알 수 있다. 의미·담화의 경우 도출된

주제가 고르게 연구되고 있었는데, 공용어로서의 영어(EFL)의 의사소통(TP2)과 리터러시 실천 및 담론(TP8)은 응용언어학과 연관성이 강해 보인다. 음성·음운론의 경우 음운론보다는 음성학 주제의 연구가 지난 10년간 많이 출판되었다.

5. 논의 및 결론

본 연구는 영어학 분야의 대표적인 학술지인 《영어학》의 지난 10년 동안의 연구 동향을 파악하기 위해 2018년부터 2023년까지 게재된 논문 456개에 대한 텍스트 마이닝을 실시하였다. 빈도 및 연관어 분석, 연도별 TF-IDF 분석, LDA 토픽모델링을 실시하여 도출한 결과는 다음과 같다.

첫째, 지난 10년간 《영어학》에 게재된 영어학 연구 분야 중 가장 주도적인 분야는 응용언어학·영어교육이다. 빈도수가 높은 단어와 이와 연관된 키워드 모두 응용언어학·영어교육과 관련성이 높았으며, 토픽모델링을 통해 도출된 주제 수와 관련 논문 수에서도 과반수를 차지하였다. 영어학 분야에서 응용언어학·영어교육 관련 연구가 많이 진행된 이유는 다양하겠지만, 무엇보다 한국 사회에서의 외국어 혹은 제 2언어의 역할을 하는 영어의 지위와 관련되어 보인다. 예를 들어, 교수법, 교육과정, 교사교육과 같은 영어교육 분야 뿐만 아니라, 일반언어학과 관련성이 높은 언어 형식의 발달과 처리에 관한 연구도 제 2언어 맥락에서 수행되면 응용언어학의 한 분야인 제 2언어습득 연구로 분류될 수 있다. 즉, 영어학이 영어를 분석의 대상으로 삼는다면, 영어가 외국어 혹은 제 2언어인 한국의 언어 맥락에서 응용언어학 및 영어교육 연구가 많이 수행될 수 밖에 없다.

둘째, 응용언어학·영어교육에서 주도적인 연구 분야는 언어 기능과 관련해서는 쓰기, 언어 형식에서는 어휘 분야로 가늠된다. 특히, 쓰기의 경우 EAP(English for Academic Purposes) 맥락에서 많이 진행된 것으로 보인다. ‘write’는 빈도가 상당히 높은 단어(5위)였을 뿐만 아니라 학술 글쓰기와 관련된 ‘argumentative’, ‘academic’, ‘genre’와 같은 단어와 관련성이 높았다. 토픽모델링의 결과에서도 ‘학술 글쓰기’와 ‘대학 영어 수업’이 주제로 도출되었으며, 관련된 논문수도 상대적으로 많다. 어휘의 경우, 빈도 분석에서 ‘word’(5위)와 ‘vocabulary’(19위)가 추출되었으며, 토픽모델링의 결과도 ‘교과서 어휘분석’, ‘어휘교수/학습’이라는 주제가 도출되었음을 보여준다.

셋째, 일반언어학과 관련된 연구에서는 통사론 분야가 주도적이다. 빈도분석에서 ‘syntax’라는 학문 영역을 특정하는 단어가 추출되었음을 물론, 토픽모델링 결과에서도 주제와 해당 논문수가 많았다. 무엇보다 통사론 분야에서 추출된 주제는 특정 통사적 현상으로 그 내용이 상대적으로 구체적이었다.

넷째, 영어학 학문 영역에서 지속적으로 다양한 연구 주제가 탐색되고 있다. 토픽모델링의 결과 40개의 주제가 고르게 분포되어 있음을 확인할 수 있었고, 연도별 TF-IDF에서 추출된 키워드와 관련 연구 논문을 통해서도 다양한 연구 주제와 연구의 독창성을 파악할 수 있었다. 이는 최근 인문학 분야의 위기 속에서도 영어학 분야 연구자들이 새로운 연구 주제를 발굴하며 학문적 토양을 넓혀가고 있음을 시사한다.

다섯째, 코퍼스언어학, 심리언어학, 전산언어학, 사회언어학 등 새로운 영어학 분야가 확인되고 있다. 연구 주제이자 방법으로서 코퍼스의 중요성은 빈도 분석과 토픽 모델링의 결과에서 확인할 수 있다. ‘corpus’라는 단어가 고빈도 단어로 추출되었을 뿐만 아니라, 순위도 높으며, 토픽모델링에서도 ‘코퍼스 분석’(TP33)이 추출되었다. 언어처리를 다루는 심리언어학의 경우, ‘process’라는 단어의 고빈도, ‘언어 처리’(TP3) 토픽 추출 등을 통해 주요한 연구 분야로 자리매김하고 있음을 알 수 있다. 전산언어학과 사회언어학의 경우, 빈도수와 토픽모델링을 통해서 확인되지는 않지만, TF-IDF 분석 결과를 통해, 새롭게 부상하고 있는 영역임을 유추할 수 있다. 2019 년 ‘cohmatrix’라는 키워드 추출, 2023 년에 추출된 ‘suicide’가 자살 유서 감지 알고리즘 개발에 대한 연구(Lee 2023)라는 점을 통해 전산언어학이 영어학 분야에서도 시도되고 있음을 알 수 있다. 사회언어학도 2018 년 ‘neoliberal’, 2021 년 ‘tagalog’, 2023 년 ‘translingualism’ 등을 통해, 비판사회언어학, 다중언어주의, 횡단언어학기 등 사회언어학 주제가 등장하고 있음을 확인할 수 있다.

여섯째, 영어학의 연구 주제도 사회 변화에 영향을 받고 있다. 4 차 산업혁명과 인공지능을 비롯한 과학기술에 대한 사회적 관심을 반영하여, 2019 년 ‘VR’, 2021 년 ‘chatbot’, 2023 년 ‘ChatGPT’가 TF-IDF 값이 높은 키워드로 추출되었다. 특히, 2021 년에는 챗봇과 관련된 연구가 집중적으로 출판되었다(Sin et al. 2021, Kim et al. 2021, Yang and Kim 2021, 권은영 2021). 또한, 코로나 19가 장기화되면서 2022년에는 코로나 19 시기 영어학습(양인영 2022, 주경진, 이경량 2022), 교육정책(Jang et al. 2022), 담론(권연진 2022)에 대한 연구도 나타났다.

마지막으로 연구방법과 관련하여, 빈도수와 연관어 분석을 통해서 정성연구보다는 정량연구가 우세함을 유추할 수 있고, 대표적인 연구 설계는 집단 간 비교 연구와 효과 측정 연구임을 알 수 있다. 이는 L1 화자와 L2 화자를 비교하거나 영어능숙도에 따른 집단 간 차이를 검증하는 하는 경우가 많은 제 2 언어습득 연구와 교수학습방법의 효과성을 검증하려는 영어교육연구가 상대적으로 많이 출판된 것과 연관이 있는 듯 하다.

본 연구는 텍스트 마이닝을 통해 《영어학》의 연구동향을 분석하였고, 이를 위해, 빈도 및 연관어 분석, TF-IDF 분석과 LDA 기반 토픽 모델링을 수행하였다. 456 편의 논문 초록에서 추출한 단어를 기반으로 통계적인 계산을 수행해 전반적인 연구 경향을 추측한 것으로 구체적인 연구 내용과 특징을 반영한 것은 아니다. 그 결과 몇몇 분석 결과는 다소 직관적이고 일반적인 것이 있다. 예를 들어, 영어학 분야에서 응용언어학과 영어교육 관련 연구가, 일반언어학 분야에서는 통사론 연구가 많이 수행되고 있다는 발견은 한국의 영어학 연구 지형을 고려했을 때 충분히 예상이 가능하다. 하지만 본 연구는 이를 계량정보학 분석 방법을 통해 실증적으로 파악했다는 데 의미가 있다. 또한, 본 연구는 영어학 세부 분야의 연구 동향도 파악했지만, 본 연구의 범위와 영어학 세부 영역 내 연구 주제 및 관행의 이질성을 고려했을 때, 보다 자세한 특징에 대해서는 후속 연구가 필요하다. 가령, 《영어학》 학술지에 출판된 통사론 분야의 연구를 추출하여 연구 동향을 분석한다면, 관련 세부 주제의 의미를 보다 자세히 파악하고 이를 언어학 분야의 통사론 연구 동향과 비교할 수 있을 것이다. 마지막으로 새로운 학문 영역의 등장과 영어학 내 학제간 연구의 증가로 영어학의 세부 학문영역을 분류하는 작업에 자의적인 부분이 있다. 영어학과 같이 다양한 학문 영역을 포괄하는 학술지의 경우 기존 학문적 경계를 넘어 연구동향을 파악할 수 있는 방법을 모색해야한다. 이러한 의미에서 후속 연구에서 최근 주목받고 있는 체계적

문헌고찰(systematic review)도 시도해볼 수 있을 것이다(Zawacki-Richter et al. 2020). 이러한 본 연구의 한계를 고려했을 때, 무엇보다 본 연구에서 계량화된 결과에서 나타나는 특정 영역의 우세함이 질적인 학문적 성과로 해석되는 것은 경계해야 할 필요가 있다.

이러한 한계에도 불구하고 《영어학》의 지난 10 년간 연구 동향 파악을 통해, 영어학 분야의 발전을 확인할 수 있었다. 이러한 발전이 지속하기 위해서는 본 연구에서 나타난 바와 같이 사회 변화 속에서 영어학의 연구의 필요성을 환기하며, 새로운 연구 주제와 방법에 대한 개방성이 필요하다. 2020 년 SCOPUS 등재 이후 양적인 면에서 향상이 이루어진 만큼, 영어학 분야에 대한 정체성에 대한 고민은 《영어학》이 학술지로서 질적인 향상을 가져오는 데 기여할 것이다. 이러한 측면에서 《영어학》이 추구하는 “영어 자체에 대한 관심과 영어의 실증적인 면에 대한 연구”가 의미하는 바를 지금까지의 성과를 바탕으로 발전시키되 새로운 주제와 방법의 등장에 대해 열린 자세로 지속적으로 탐구하고 추구할 필요가 있다.

References

- 권민재(Kwon, M.). 2024. 언어 네트워크 분석을 활용한 독어학 연구 동향 분석 - 최근 25 년간 발표된 독어학 관련 논문을 중심으로(Analyse der forschungstendenzen in der germanistischen linguistik in Korea anhand der sprachlichen netzwerkanalyse). 《독일언어문학》(*Zeitschrift für Deutsche Sprache und Literatur*) 103, 1-26.
- 권연진(Kwon, Y.). 2022. 정치담론 상에서 코로나 19 은유의 개념화 양상(The conceptualization of Covid-19 metaphor in political discourse). 《영어학》(*Korean Journal of English Language and Linguistics*) 22, 871-890.
- 권은영(Kwon, E.). 2020. 파이썬과 텍스트 마이닝을 활용한 ESP 연구 동향 분석. 《외국어교육》(*Foreign Languages Education*) 27-2, 111-139.
- 권은영(Kwon, E.). 2021. AI 활용 영어 교수 및 학습 연구 동향(Research trends in AI-based English language teaching and learning). 《영어학》(*Korean Journal of English Language and Linguistics*) 21, 1313-1337.
- 김미경(Kim, M.). 2023. ChatGPT 와 Google Bard 를 활용한 Critical-PBLL 중심 대학영어 개발과 적용(Towards a Critical-PBLL utilizing ChatGPT and google bard within college English education). 《영어학》(*Korean Journal of English Language and Linguistics*) 23, 741-767.
- 남재우, 김영준(Nam, J. and Y. Kim). 2023. 텍스트 마이닝을 이용한 인공지능 활용 신약 개발 연구 동향 분석(Analysis of research trends in new drug development with artificial intelligence using text mining). 《생명과학회지》(*Korean Journal of Life Science*) 33-8, 663-679.
- 박상희, 전지현(Park, S. and J. Jeon). 2019. 한국과 미국 직장인의 영어 요청 이메일 비교 연구(A study on English request emails by Korean and American office professionals). 《영어학》(*Korean Journal of English Language and Linguistics*) 19-1, 142-164.
- 박신애(Park, S.). 2022. 텍스트 마이닝을 활용한 영어 음운의 국내 연구 동향 분석(An analysis of domestic study trends in English phonology using text mining). 《음성음운형태론연구》(*Studies in Phonetics, Phonology, and Morphology*) 28-1, 83-110.

- 박은희(Park, E.). 2021a. LDA 토픽 모델링을 활용한 대학 교양영어 연구동향 분석(Research trends in college English education in Korea – A topic analysis using LDA topic modeling). *《교양교육연구》(Korean Journal of General Education)* 15-5, 169-183.
- 박은희(Park, E.). 2021b. 텍스트 마이닝을 활용한 영어 쓰기교육 연구동향 분석(Topic analysis in EFL writing in Korea using text mining). *《응용언어학》(Korean Journal of Applied Linguistics)* 37-3, 95-122.
- 성민창, 김기택(Seong, M. and K. Kim). 2021. 우리나라 초등영어교육에 적합한 팝송 탐색: 언어 수준과 내용 분석(Pop songs for primary English education in South Korea: An analysis of language level and content). *《영어학》(Korean Journal of English Language and Linguistics)* 21, 450-471.
- 신동일(Sin, D.). 2018. 글로벌 인재와 영어능력에 관한 담론적 실천과 신자유주의 주체성의 이해(Understanding discursive practice and neoliberal subjectivities of global talents and their English language competence). *《영어학》(Korean Journal of English Language and Linguistics)* 18-3, 349-380.
- 신승훈(Sin, S.). 2014. 치경비음을 선행하는 치경폐쇄음의 발화양상: 벅아이 코퍼스의 미국영어를 중심으로(Realizations of alveolar stops before alveolar nasals in the Buckeye corpus of conversational speech). *《영어학》(Korean Journal of English Language and Linguistics)* 14-2, 279-300.
- 신유선·김양희(Sin, Y. and Y. Kim). 2020. 키워드 네트워크 분석을 활용한 영어교육관련 연구 동향 분석: 2000-2019 년도까지 「영어교육연구」에 게재된 논문을 중심으로(Research and trends in English language education from 1990 to 2019: A keyword analysis of published articles in English Language Teaching). *《영어교육연구》(English Language Education)* 32-4, 205-224.
- 심영숙(Sim, Y.). 2020. 세계영어에 관한 국내 연구 동향 분석(Trend analysis of world Englishes research published in Korean domestic journals). *《응용언어학》(Korean Journal of Applied Linguistics)* 36-3, 99-125.
- 양인영(Yang, I.). 2022. 고등학교 영어 기초학력 미달에 영향을 미치는 학생 특성 요인: COVID-19 전후 비교(Effect of student characteristics on the high school students' below-basic achievement level in English: Focusing on the changes after COVID-19 outbreak). *《영어학》(Korean Journal of English Language and Linguistics)* 22, 208-228.
- 원용국·김영우(Won, Y. and Y. Kim). 2021. 토픽 모델링을 활용한 한국 영어교육 학술지에 나타난 연구동향 분석(Analysis of research trends in Korean English education journals using topic modeling). *《한국콘텐츠학회논문지》(The Journal of the Korea Contents Association)* 21-4, 50-59
- 이은주(Lee, E.). 2008. 영어교육과 응용언어학 분야에서 수행된 코퍼스 기반 연구의 분석(An analysis of corpus-based research on TEFL and applied linguistics). *《영어교육》(English Teaching)* 63-2, 283-306.
- 조선, 신동일(Cho, S. and D. Sin). 2014. 전화영어에 나타난 대화의 특성 연구: 상호작용 담화모형의 관점에서(Investigating characteristics of telephone English conversation: from the analytic framework of an interactional discourse model). *《영어학》(Korean Journal of English Language and Linguistics)* 14-4, 603-631.
- 주경진, 이경량(Ju, K. and K. Lee). 2022. 포스트코로나 시대 대학생들의 영어 의사소통역량 함양과 자기주도적 학습 능력 향상 방안 모색 연구(An exploratory study to improve English

- communicative competencies and self-directed learning of university students in the post COVID-19 era). 《영어학》(*Korean Journal of English Language and Linguistics*) 22, 1220-1235.
- 차경환·김영규(Cha, K. and Y, Kim). 2010. 한국응용언어학회 학회지 게재 논문(1983-2008)에 나타난 연구 경향 분석(A trend analysis of research published in the journal of the applied linguistics association of Korea (1983-2008)). 《응용언어학》(*Korean Journal of Applied Linguistics*) 26-1, 91-110.
- 황보영식(Hwang, B.). 2017. 미국영어에서 [u]-약화와 [w]-삽입([u]-reduction and [w]-epenthesis in American English). 《영어학》(*Korean Journal of English Language and Linguistics*) 17-2, 351-378.
- Aizawa, A. 2003. An information-theoretic perspective of TF-IDF measures. *Information Processing & Management* 39(1), 45-65.
- Akoglu, H. 2018. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine* 18(3), 91-93.
- Anne Kao and R. P. Stephen. 2007. *Natural Language Processing and Text Mining*. London: Springer-Verlag.
- Arun, R., S. Suresh, C. E. V. Madhavan and M. N. Murty. 2010. On finding the natural number of topics with latent dirichlet allocation: Some observations. *Lecture Notes in Artificial Intelligence* 6118, 391-402.
- Aurora, P., B. Rafael and R. José. 2007. Topic discovery based on text mining techniques. *Information Processing & Management* 43(3), 752-768.
- Blei, D., L. Carin and D. Dunson. 2010. Probabilistic topic models: A focus on graphical model design and applications to document and image analysis. *IEEE Signal Processing Magazine* 27, 55-65.
- Bulbul Gupta, M. Tabish, S. S. Sohail and D. Ø. Madsen. 2023 ChatGPT: A brief narrative review. *Cogent Business & Management* 10(3).
- Cao, J., T. Xia, J. Li, Y. Zhang and S. Tang. 2009. Density-based method for adaptive LDA model selection. *Neurocomputing* 72, 1775-1781.
- Chin, H., G. Lima, M. Shin, A. Zhunis, C. Cha, J. Choi and M. Cha. 2023. User-chatbot conversations during the COVID-19 Pandemic: Study based on topic modeling and sentiment analysis. *Journal of Medical Internet Research* 25, e40922.
- Cho, S. 2019. Does tutor feedback make a difference? Focusing on EFL Korean student writing. *Korean Journal of English Language and Linguistics* 19(3), 432-451.
- Cho, S., D. Kim and C. Baek. 2020. A comparison of a novice writing tutor with an experienced tutor: Focusing on changes in their interactions with tutees. *Korean Journal of English Language and Linguistics* 20, 829-850.
- Choi, J. and H. Yun. 2022. Effects of part and whole learning on the learning of L2 words and idioms. *Korean Journal of English Language and Linguistics* 22, 441-457.
- Choi, Y., H. Ahn and J. Lee. 2018. Anaphora resolution strategies in L2 reading. *Korean Journal of English Language and Linguistics* 18(4), 395-422.
- Crosthwaite, P., S. Ningrum and I. Lee. 2022. Research trends in L2 written corrective feedback: A bibliometric analysis of three decades of Scopus-indexed research on L2 WCF. *Journal of Second Language Writing* 58, 100934.
- Das, M., S. Kamalanathan and P.J.A. Alphonse. 2023. A comparative study on TF-IDF feature weighting method and its analysis using unstructured dataset. In *Proceedings of the COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems*.
- Deveaud, R., E. SanJuan and P. Bellot. 2014. Accurate and effective latent concept modeling for ad hoc information retrieval. *Document Mumérique* 17, 61-84.

- Eroğlu, Y. and S. Seçkiner. 2016. Trend topic analysis for wind energy researches: A data mining approach using text mining. *Journal of Technology Innovations in Renewable Energy* 5, 44-58.
- Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth and R. Uthurasamy. 1996. *Advances in Knowledge Discovery and Data Mining*. CA: AAAI/MIT Press.
- Feldman, R. and I. Dagan. 1995. Knowledge discovery in textual databases (KDT). *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*. AAAI Press, 112-117.
- Field, A. 2009. *Discovering Statistics Using SPSS*. London: Sage.
- Goksu, I., E. Ozkaya and A. Gunduz. 2022. The content analysis and bibliometric mapping of CALL journal. *Computer Assisted Language Learning* 35(8), 2018-2048.
- Griffiths, T. L. and M. Steyvers. 2004. Finding scientific topics. In *Proceedings of the National Academy of Sciences* 101(Suppl 1), 5228-5235.
- Gupta, B., T. Mufti, S. S. Sohail and D. Ø. Madsen. 2023. ChatGPT: A brief narrative review. *Cogent Business & Management* 10, 3.
- Han, S. and Y. Lee. 2016. Developing a collocation test for Korean EFL learners: Test validation and item format effect. *Korean Journal of English Language and Linguistics* 16(2), 221-243.
- Harris, Z. S. 1954. Distributional structure. *Word* 10, 146-162.
- Hyland, K. and F. K. Jiang. 2021. A bibliometric study of EAP research: Who is doing what, where and when? *Journal of English for Academic Purposes* 49, 100929.
- Jang, I. H. Kim and E. Choi. 2022. Exploring parental satisfaction with public English education during the COVID-19 pandemic in South Korea. *Korean Journal of English Language and Linguistics* 22, 377-395.
- Jin, H. 2023. Lexical frames and errors in the use of English definite article in L2 academic writing: A case of English placement test. *Korean Journal of English Language and Linguistics* 23, 324-341.
- Kang, D. 2022a. Effects of single translation marginal & multiple-choice glosses on L2 academic vocabulary learning. *Korean Journal of English Language and Linguistics* 22, 1417-1427.
- Kang, D. 2022b. An exploratory study of tasks & glosses on L2 academic vocabulary learning. *Korean Journal of English Language and Linguistics* 22, 1300-1312.
- Kang, D. 2022c. Effects of multiple-choice glosses & frequency on L2 academic vocabulary learning in Korea. *Korean Journal of English Language and Linguistics* 22, 695-705.
- Kang, H. and C. Lee. 2020. Effects of focus on form instruction through listening in blended learning on the development of grammar and listening skills. *Korean Journal of English Language and Linguistics* 20, 662-691.
- Kim, H., Y. Cha and N. Kim. 2021. Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics* 21, 712-734.
- Kim, M, K. Ohk and C. Moon. 2017. Trend analysis by using text mining of journal articles regarding consumer policy. *New Physics: Sae Mulli* 67, 555-561.
- Kim, Y. 2016. Tough-constructions vs non-tough constructions. *Korean Journal of English Language and Linguistics* 16(3), 493-512.
- Kobayashi, V., S. T. Mol, H. A. Berkers, G. Kismihók and D. N. D. Hartog. 2018. Text classification for organizational researchers: A tutorial. *Organizational Research Methods* 21(3), 766-799.
- Kobayashi, V., S. T. Mol, H. A. Berkers, G. Kismihók and D. N. D. Hartog. 2018. Text classification for organizational researchers: a tutorial. *Organizational Research Methods* 21(3), 766-799.
- Kwon, E. and E. Park. 2015. Learners' awareness of comma usage and its effect on two types of writing tasks. *Korean Journal of English Language and Linguistics* 15(4), 713-738.

- Lawani, S. 1981. Bibliometrics: Its theoretical foundations, methods and applications. *Libri*, 31, 294-315.
- Lee, a. 2019. Using virtual reality to test academic listening proficiency. *Korean Journal of English Language and Linguistics* 19(4), 688-712.
- Lee, J. 2014. Student annotations triggered by teacher feedback in L2 business writing. *Korean Journal of English Language and Linguistics* 14(4), 655-684.
- Lee, S., J. Baker, J. Song and J. C. Wetherbe. 2010. An empirical comparison of four text mining methods. *43rd Hawaii International Conference on System Sciences* 1-10.
- Lee, Y. 2023. Detecting suicide notes with the probability of positive sentiment and interquartile range. *Korean Journal of English Language and Linguistics* 23, 431-447.
- Lei, L. and D. Liu. 2019a. Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics* 40(3), 540-561.
- Lei, L. and D. Liu. 2019b. The research trends and contributions of *System*'s publications over the past four decades (1973-2017): A bibliometric analysis. *System* 80, 1-13.
- Li, W. 2018. Translanguaging as a practical theory of language. *Applied Linguistics* 39(1), 9-30.
- Limna, P., T. Kraiwanit, T. K. Jangjarat, P. Klayklung, and P. Chocksathaporn. 2023. The use of ChatGPT in the digital era: Perspectives on chatbot implementation. *Journal of Applied Learning and Teaching* 6(1), 64-74.
- Lin, Z. and L. Lei. 2020. The research trends of multilingualism in applied linguistics and education (2000-2019): A bibliometric analysis. *Sustainability* 12(15), 6058.
- Mackey, A. and S. M. Gass. 2022. *Second Language Research: Methodology and Design* (3rd ed.). New York: Routledge.
- Park, H. 2015. The effects of cognitive complexity on summary quality by Korean college students. *Korean Journal of English Language and Linguistics* 15(3), 463-491.
- Park, J. S.-Y. and A. Lo. 2012. Transnational South Korea as a site for a sociolinguistics of globalization: Markets, timescales, neoliberalism. *Journal of Sociolinguistics* 16(2), 147-164.
- Plonsky, L. 2023. Introduction to the special issue on Introducing bibliometrics in applied linguistics. *Studies in Second Language Learning and Teaching* 13(4), 721-726.
- Pritchard, A. 1969. Statistical bibliography or bibliometrics? *Journal of Documentation* 25(4), 348-349.
- Schober, P., C. Boer and L. A. Schwarte. 2018. Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia* 126(6), 1763-1768.
- Sebastiani, F. 2002. Machine learning in automated text categorization. *ACM Computing Surveys* 34, 1-47.
- Sin, D., H. Kim, J. Lee and H. Yang. 2021. Exploring the use of an artificial intelligence chatbot as second language conversation partners. *Korean Journal of English Language and Linguistics* 21, 375-391.
- Wiedemann, G. 2013. Opening up to big data: computer-assisted analysis of textual data in social sciences. *Historical Social Research/Historische Sozialforschung* 38(4), 332-357.
- Yang, D. 2017. The Determinism Issue in the minimalist theory: The derivational null hypothesis. *Korean Journal of English Language and Linguistics* 17(3), 429-448.
- Yang, I. 2018. Pause – Tonic Stress Interaction in English L2 Speech of Korean Talkers with Different Proficiency Levels. *Korean Journal of English Language and Linguistics* 18(1), 1-29.
- Yang, H. and H. Kim. 2021. Development and application of AI chatbot for cabin crews. *Korean Journal of English Language and Linguistics* 21, 1085-1104.
- Yang, X. D. and B. Jia. 2010. Vector space model based on Lucene index and TF-IDF weighting algorithm. *Proceedings of 2010 Asia-Pacific Youth Conference on Communication* 1-2, 20-23.

- Yu, M. 2023. The Effect of referential forms on incremental processing of long-distance dependencies: Evidence from English cleft sentences. *Korean Journal of English Language and Linguistics* 23, 913-934.
- Zawacki-Richter, O., Kerres, M. Bedenlier, S. Bond, M and Buntins. K. (2020). *Systematic Reviews in Educational Research: Methodology, Perspectives and Application*. Wiedbaden: Springer Nature.
- Zhang, W., T. Yoshida and X. Tang. 2011. A comparative study of TF* IDF, LSI and multi-words for text classification. *Expert Systems with Applications* 38(3), 2758-2765.
- Zhang, X. 2020. A bibliometric analysis of second language acquisition between 1997 and 2018. *Studies in Second Language Acquisition* 42(1), 199-222.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary