

## 고등학교 영어 듣기 평가 영역에서의 코퍼스 기반 어휘다발 분석

박소연 (이화여자대학교)  
이은주 (이화여자대학교)

Received: 13 March 2025  
Revised: 17 April 2025  
Accepted: 25 April 2025

Park, Soyoun, & Lee, Eun-Joo. (2025). A Corpus-based Analysis of Lexical Bundles in the Listening Section of High School English Assessments. *Modern English Education*, 26, 213-229.

### Keywords

Lexical bundles, multi-word expressions, formulaic sequences, listening assessment, high school English 어휘다발, 단어회표현, 정형어구, 듣기평가, 고등학교 영어

### Soyoun Park

(First author)  
Department of English Education  
Ewha Womans University  
[soyouni94@ewha.ac.kr](mailto:soyouni94@ewha.ac.kr)

### Eun-Joo Lee

(Corresponding author)  
Professor  
Department of English Education  
Ewha Womans University  
[eunlee@ewha.ac.kr](mailto:eunlee@ewha.ac.kr)

\*본 논문은 제1저자의 석사 학위논문을 수정 보완한 것입니다.

### Abstract

This study investigates the use of lexical bundles in the listening section of high school English assessments. To better understand authentic language use, facilitate learners' vocabulary development, and inform effective pedagogical strategies and test design, this study analyzed frequency, structural, and functional differences among lexical bundles in English listening sections of the CSAT, Academic Achievement Tests, Mock Exams, and a native English reference corpus (i.e., the American National Corpus). Results indicated that listening assessments contained more lexical bundles than the ANC. Structural analysis showed that phrasal bundles were the most frequently used in both corpora, with verb phrases being the second most common in listening assessments, whereas noun phrases were the second most common in the ANC. Functional analysis revealed that stance bundles were most frequently employed in both corpora. Special conversational functions were the second most frequent in listening assessments, while they were the least frequent in the ANC. Additionally, referential expressions and discourse organizers occurred less frequently than other functions. Both corpora thus exhibited characteristics of a spoken register. These findings underscore the importance of context-based learning that focuses on multi-word expressions and lexical bundles to enhance learners' vocabulary skills.

## 서론

영어를 배우는 학습자들의 궁극적인 목표는 의사소통 능력을 효율적으로 증진시키는 것이다. 이에 따라 영어교육 현장과 연구 분야에서는 학습자들의 의사소통 능력 향상에 초점을 맞추고 이를 향상시키기 위한 여러가지 방안을

제시하고 있으며 특히 최근에는 어휘학습과 언어학습의 중요성을 강조하고 있다(Boers et al., 2014; Lewis, 2000; McCarthy & O'Dell, 2005). 예를 들면, Lewis(1993)는 언어의 가장 기본 단위는 어휘라고 주장하며 제2 언어를 학습하는데 있어 어휘 습득의 중요성을 역설하였다. 즉, 제2언어 습득에서 어휘적 접근 방식을 제시하면서 이는 어휘 구문(lexical phrase)을 하나의 덩어리(chunk)로 이해하고 생성할 수 있게 하는 것이라고 설명하였다. 한편, Granger(1998)는 어휘다발을 텍스트에서 빈번하게 함께 나타나는 세 개 이상의 연속된 단어들로 정의하면서 언어 사용에 중요한 역할을 한다고 주장하였다. 즉, 어휘다발은 학습자들에게 정확한 의미 전달을 가능케하며, 미리 조합된 표현을 사용함으로써 의사소통시 언어를 생성하는 시간을 줄여주고 유창성을 높여준다는 것이다. 동일한 연구분야에서 Biber와 Barbieri(2007)는 어휘다발을 사용하는 것이 담화를 효율적으로 이해하고 발화하는데 필수 요소라고 설명하였다. 이처럼 어휘와 어휘다발은 언어 능력의 핵심적인 요소이며, 의사소통 중심 교육에서의 역할이 지대하다고 할 수 있다.

그러나 우리나라 영어 교육 현장에서는 어휘를 능동적이고 유창하게 사용하게 하는 지도와 학습 보다는 영어 어휘의 형태와 의미를 암기하는 학습 방법에 중점을 두고 있다. 그로 인하여 대다수의 학습자들은 많은 어휘를 알고 있음에도 불구하고 실제 의사소통 상황에서 어휘를 어떻게 결합하고 사용하는가에 대한 언어 능력이 부족한 실정이다(Hill, 2000; McCarthy, 1990). 우리나라와 같은 EFL 교육환경에서 학습자들에게 입력되는 언어는 실제 영어를 사용할 수 있게 하는 중요한 요소이므로, 학습자들에게 빈번하게 노출되고 있는 학습매체 또는 입력 자료가 학습에 미치는 영향이 지대하다고 볼 수 있다. 이러한 상황에서 대학수학능력시험은 고등학교 수업과 평가에 영향을 끼치며, 학생들에게 중요한 학습 자료로 사용된다(Kim & Lee, 2019). 또한, 대학수학능력시험과 모의평가, 전국연합학력평가는 한국 고등학생들의 학력 수준을 평가하기 위한 도구이며 교수법에도 영향을 미치는 역류효과가 크기 때문에 평가 도구임과 동시에 입력자료로서 분석 가치가 있다. 그 중 영어 듣기 평가는 학생들의 실제 의사소통 능력을 간접적으로 평가하는 중요한 도구 중의 하나라고 할 수 있다. 이에 따라서 평가 지문을 분석하고 활용하고자 하는 연구가 다수 진행되었다. 그러나 이러한 연구에서는 읽기 지문 중심의 코퍼스를 기반으로 텍스트 자료를 비교 분석하는 연구들이 대부분이며, 구어 텍스트를 분석한 연구도 진행되었으나 영어 듣기 평가에서 사용되는 어휘다발에 대한 연구는 상대적으로 부족하다.

영어교육 현장에서 점차 어휘와 어휘다발의 중요성이 강조되면서 이를 분석하는 연구가 진행되었으며, 특히 코퍼스를 활용한 연구가 수행되면서 영어 교과서, EBS와 원어민 참조 코퍼스를 활용하여 어휘다발의 구조적, 기능적 특징을 비교 분석하는 시도들이 있었다. 그러나, 고등학교 영어 평가를 중심으로 분석한 연구는 아직 부족한 실정이며, 수능이나 영어교과서를 바탕으로 한 대부분의 연구들이 읽기 지문 분석 연구에 중점을 두고 있고 듣기 영역에 나타나는 어휘 및 어휘다발을 분석하는 연구는 매우 부족한 실정이다. 이에 본 연구에서는 2013학년도부터 2024학년도까지의 수능, 모의평가, 학력평가의 영어 듣기 평가 텍스트에서 나타난 어휘다발을 추출하여 구조적, 기능적 특징을 파악하고 원어민 코퍼스인 American National Corpus(ANC)와 비교 분석하고자 한다.

## 이론적 배경

### 코퍼스(Corpus)와 코퍼스 언어학

코퍼스는 문자 또는 음성 언어를 모은 언어 자료이며, 실제 의사소통 상황에서 언어의 다양한 특성을 파악하기 위해 언어 텍스트를 모아 놓은 것으로 정의된다(Sinclair, 1991). 컴퓨터 기술의 발달로 코퍼스 구축이 가능해지고 다수의 언어 분석 프로그램이 개발되면서 언어 자료를 체계적으로 분석할 수 있게 되었다. 코퍼스 기반으로 한 분석을 통해 어휘 목록, 빈도수와 유형, 단어 조합, 분포 등 텍스트의 특징을 객관적으로 파악할 수 있게 되었으며, 실제 원어민 화자들이 사용하는 데이터를 확인하고, 구어와 문어에서 빈번하게 사용되는 표현을 파악하며, 해당 어휘와 함께 나타나는 언어를 효과적으로 학습할 수 있게 되었다.

이와 같이 코퍼스는 영어 교육 분야에서 활발히 활용되며 코퍼스를 통해 실제로 사용되는 언어 자료를 분석하여 빈도수 높은 어휘와 어휘다발을 추출할 수 있다. 코퍼스 활용 연구 분야는 크게 어휘, 문법, 담화, 교육으로 분류할 수 있는데 이 중 어휘 영역에서 가장 많은 연구가 수행되었다(E. Lee, 2009). 코퍼스 언어학은 대규모 텍스트 데이터를 분석하여 언어 사용의 실제 패턴을 파악하는 데 도움이 되므로 EAP(English for Academic Purpose)와 ESP(English for Specific Purpose)분야에서도 단어 연속체 관련 연구가 활발하게 진행되고 있다(Appel, 2022; Coxhead

et al., 2017; Liu & Chen, 2020; Staples et al., 2013; Wang & Csomay, 2024). 이처럼 코퍼스를 활용한 어휘 연구는 언어 사용의 양상을 파악하고 빈도 분석, 분포 등 다양한 특성을 객관적으로 분석할 수 있는데 그 특징은 다음과 같다. 첫째, 코퍼스는 학습자들에게 다양한 맥락에서 단어와 구문의 실제 사용을 예시로 제공하며 학습자들이 어휘, 관용적 표현과 언어 지식 등을 실제 맥락에서 학습할 수 있게 해준다(Nation, 2001). 둘째, 코퍼스는 교재 개발에 중요한 자료로 활용되며 의사소통 중심의 교재 제작 시, 원어민의 실제 언어 사용의 빈도와 패턴을 반영한다. 셋째, 코퍼스의 통계적 분석 기능을 통해 고빈도 어휘 및 어구를 파악하여 효과적인 학습 자료를 제공할 수 있다. 마지막으로, 코퍼스는 중점적으로 가르쳐야 할 내용을 선별하는데 중요한 역할을 하며 원어민들의 실제 사용 빈도를 기반으로 학습 내용을 선정할 수 있게 한다(Biber, 1998; Nation, 2001). 이와 같이 코퍼스는 언어 사용의 빈도수와 예문을 많이 제공함으로써 언어교육 관련 교재 개발에 중요한 역할을 하며 언어 교육의 질적 향상에도 기여한다고 할 수 있다(E. Lee, 2009).

## 어휘다발의 정의 및 특성

어휘다발(lexical bundle)은 Biber 외 4인(1999)이 처음 사용한 용어로 텍스트에서 빈번하게 나타나는 단어의 연속체이다. 즉, 어휘다발은 언어(collocation)에서 확장된 개념으로, 함께 자주 나타나는 연속된 단어를 의미한다. 어휘 다발의 정의인 단어 연속체라는 의미를 지닌 유사한 용어로는 어휘구문(lexical phrases), 정형어구(formulaic sequences), 다어휘 표현(multiword expressions), 고정화된 패턴(prefabricated patterns), 어휘군집(clusters), 어휘덩치(chunks), 엔그램(N-grams) 등이 있다.

어휘다발은 빈도, 어휘 길이, 분포와 같은 요소로 분류할 수 있는데 그 기준은 다음과 같다. 첫째, 어휘 다발의 최저 빈도수는 일반적으로 코퍼스 규모 100만 단어당 10회에서 40회, 20회에서 40회의 출현을 기준으로 한다(Biber et al., 2004; Chen & Baker, 2010; Hyland, 2008). 반면 구어 코퍼스와 같은 소규모 코퍼스의 경우, 2회에서 10회 최저빈도(cut off frequency)가 자주 사용되며, 코퍼스의 크기가 작은 경우에는 최저빈도수를 낮게 정한다(Kim & Lee, 2019). 둘째, 어휘 다발의 어휘의 길이는 일반적으로 2에서 6 단어까지 분석하며 3 단어는 4 단어에 부분적으로 포함되는 경향이 있어 중복으로 분석되는 경우가 있다. 따라서, 많은 연구에서는 4단어 어휘다발에 중점을 두어 분석하고 있다. 이는 4단어는 5단어보다 빈도수가 높고, 다양한 구조와 기능 분석이 가능하기 때문이다(Biber et al., 2004; Cortes, 2004; Hyland, 2008; Simpson-Vlach & Ellis, 2010). 셋째, 어휘다발의 분포를 살펴보면, 어휘다발은 일반적으로 최소 3-5개의 텍스트에서 출현해야 하며 전체 텍스트의 10%에서 나타나야 한다(Biber & Barbieri, 2007). 이러한 기준들은 관용구와 언어와 같이 어휘다발의 다른 유형의 다단어 표현과 구별할 수 있다.

## 선행연구

최근 코퍼스 언어학의 발달에 따라 코퍼스를 이용한 응용언어학적 연구가 활발하게 수행되고 있으며, 특히 어휘다발을 통한 어휘 학습에 관한 관심이 높아지면서 관련 연구들이 다양하게 수행되고 있다. 코퍼스 기반의 어휘 및 어휘다발에 대한 선행연구들을 살펴보면 다음과 같다. Kwon(2021)은 대학생들의 영어 말하기 과제로 작성한 자기소개 대본에 나타난 어휘다발을 분석하였는데, Lancsbox 6.0을 사용하여 4단어 어휘다발을 추출한 후, 온라인 설문조사와 인터뷰를 실시하였다. 연구 결과, 학습자들은 교수자가 제공한 자료에 많이 의존하였지만 본인의 아이디어를 첨가해서 대본을 작성하였다. 학생들은 참고자료가 과제 수행에 도움이 되었다고 보고하였으며 제공된 자료가 학습 효율을 높여주고 다양한 예시가 제시되어야 학습의 효율성을 높일 수 있다고 설명하였다.

Shin 외 3인(2018)은 BNC-COCA 25000 Range 프로그램과 COCA\_MWU20 ColloGram을 사용하여 한국인 학습자와 영어 원어민의 구어와 문어 코퍼스에 나타난 어휘와 다어휘 표현을 비교하였다. 분석 결과 학습자들은 구어와 문어 사이의 어휘 및 다어휘 표현 사용의 차이가 상대적으로 적었는데, 이는 학습자들의 제한된 언어 능숙도로 인해 문어체에서 구어체와 비슷한 표현을 사용하는 경향이 있기 때문이라고 분석하였다. 또한, 개별 어휘는 상위 5,000개 단어, 다어휘 표현의 경우 상위 2,500개 표현이 높은 사용 빈도로 나타났다. 특히 최상위 500개의 다어휘 표현이 압도적으로 높은 빈도로 사용되었다. 또다른 결과로는 구어에서는 제한된 표현이 반복적으로 사용되었고 문어에서는 상대적으로 다양한 표현이 사용되었다. 특히 다어휘 표현은 구어에서 더 빈번하게 사용되는 것으로 나타나 의사소통 상황에서 인지적 부담을 줄이기 위한 것으로 나타났다. 마지막으로,

한국인 학습자들은 빈도와 실용성이 낮은 표현을 학습하는 데 시간과 노력을 쏟는 경향이 있다고 밝혔다.

통번역 연구에서도 어휘 다발의 특성을 분석한 연구가 수행되었는데, Wu(2021)의 연구는 유럽 의회(EP)의 통역사, 번역사, 그리고 의원들의 4단어 어휘다발 사용의 유사점과 차이점을 코퍼스를 기반으로 분석하였다. EP회의에서 영어 통역, 번역, 연설을 각각 분류하여 구조적, 기능적 분석을 수행하였는데, 구조적 분석 결과, 세 레지스터에서 모두 명사구와 전치사구 어휘다발이 가장 빈번하게 사용되며 동사구 어휘다발이 가장 낮은 비율을 차지하였다. 기능적 분석 결과, 번역과 연설에서 가장 빈번하게 등장한 주제 기능의 어휘다발과 지칭 어휘다발은 주로 의회 담화에서 정보를 전달하기 위한 목적으로 사용되었다. 한편, 통역에서는 양태, 지칭 기능의 어휘다발이 유사한 비율을 차지하면서 의사소통과 정보를 전달하기 위한 목적으로 나타났다.

Staples 외 3인(2013)의 연구는 언어 능력이 다른 세 집단 학생들의 TOEFL iBT 작문 영역의 응답에 나타나는 어휘다발을 분석하였다. 그 결과, 능숙도가 낮은 수준의 학습자들이 더 많은 어휘다발을 사용하였으며 모든 수준의 학습자들이 양태, 담화 조직 어휘다발을 유사한 빈도로 사용하며 지칭 어휘다발은 거의 사용하지 않았다. 또한, 화자의 입장을 표현하는 어휘다발이 담화, 지칭 표현보다 더 빈번하게 나타났으며, 시험 응시자들은 주제에 대한 입장을 표현하기 위해 양태 어휘다발을 빈번하게 사용하였다. 어휘다발의 고정성(bundle fixedness) 패턴 분석 결과 모든 수준의 학습자들의 고정된 어휘다발과 가변 어휘다발(fixed and variable slot bundles) 사용 비율이 유사한 것으로 나타났다. 가변 어휘다발에는 기능어가 주로 사용되며 비영어권 학습자들이 학술 글쓰기에서 사용하는 어휘 다발이 학술적 문체보다는 구어체에 더 가깝다는 것을 시사한다.

영어평가와 교과서 지문에 나타나는 어휘다발을 비교 분석한 연구도 수행되었는데, H. Choi(2019)는 중학교 1-3학년 교과서와 전국 영어듣기능력평가 나타난 단어 연속체 어휘다발을 교육과정에서 제시하는 의사소통 기능을 기준으로 분석하고 교과서와 듣기평가, 원어민 참조 코퍼스에서 나타나는 어휘다발의 기능적인 분포 양상 및 형태를 비교 분석하였다. 그 결과, 어휘다발의 기능별 분포가 교과서, 듣기평가, 원어민 참조 코퍼스에서 각각 다른 양상이 나타나며 이는 어휘다발이 사용되는 맥락이나 상황에 차이가 있다는 것을 알 수 있었다. 교과서, 듣기평가, 원어민 코퍼스에서는 빈도 면에서 공통적으로 양태 표현 기능이 가장 높은 비율로 나타난 반면, 지시적 표현 기능은 원어민 참조 코퍼스보다 교과서와 듣기평가에서 더 낮은 비율을 차지하였다. Kim과 Lee(2019)는 대학수학능력시험, 모의고사 영어 읽기 지문과 SAT 읽기 평가 텍스트를 코퍼스로 구축하여, 빈도수와 상호정보(MI)가 가장 높은 3, 4단어 정형어구를 추출하고, 두 코퍼스에서 나타나는 정형어구의 구조적, 기능적 특성을 살펴보았다. 연구 결과, SAT가 수능에 비해 어휘 다양성과 난이도가 높았으며 수능에서는 SAT보다 더 많은 정형어구가 나타났다. 이를 통해 수능에서 반복된 표현이 많이 사용되었음을 알 수 있었다. 구조적 분석 결과 수능에서는 명사구가 가장 빈번하게 나타났으며 SAT에서는 전치사구가 가장 많이 나타났고, 기능적 분석에서는 두 코퍼스 모두 지칭 표현이 가장 빈번하게 사용되었다고 보고하고 있다. 이를 통해, 언어 사용역에 따라 사용되는 고빈도 정형어구가 다르다는 것을 알 수 있었다. S. H. Lee(2022)는 2015 개정 교육과정 고등학교 영어 교과서의 대화문에서 나타나는 4연속 어휘다발의 출현 양상을 문법적, 기능적 범주로 분류하여 BNC baby에 나타난 어휘다발과 비교하여 특성을 보고하였다. 분석 결과, 영어교과서에서는 동사구와 절 순서로 가장 빈번하게 나타났으며, 명사구와 전치사구는 나타나지 않았다. 반면, 원어민 참조 코퍼스에서는 절, 동사구 순서로 가장 빈번하게 출현하였고 전치사구는 나타나지 않았다. 기능적 측면에서는 영어교과서와 원어민 참조 코퍼스 대부분 양태 표현이 가장 빈번하게 등장하였다. Yang(2020)은 중고등학교 영어 교과서에 나타난 3, 4단어 어휘다발의 구조적, 기능적 특성을 원어민 참조 코퍼스와 비교 분석하였다. 연구 결과, 고등학교 영어 교과서가 중학교 영어 교과서보다 더 다양한 어휘를 사용하였다. 고등학교 교과서와 원어민 참조 코퍼스에서는 명사구와 전치사구 어휘다발이 가장 빈번하게 나타나며 중학교 교과서에는 동사구 어휘다발이 높은 빈도로 사용되었다. 기능적 분석 결과, 세 코퍼스 모두 지칭적, 입장 어휘다발이 높은 빈도로 나타나고, 담화 조직 기능이 가장 낮은 빈도로 나타났다.

Park과 Lee(2021)의 연구는 고등학교 영어듣기 시험에서 나타난 단일어, 어휘다발, 탈어휘 동사 연어의 사용 양상을 분석하였다. 그 결과, 시험에 나타난 어휘가 교육과정이 설정한 기본 어휘를 벗어나 보다 더 어려운 어휘를 사용했음을 알 수 있었다. 또한, 학력평가, 고등학교 영어듣기평가에서 입장의 기능을 나타내는 어휘다발이 가장 빈번하게 나타났으나, 수능에서는 특수 담화 기능의 어휘다발이 입장 기능 어휘다발보다 더 많이 사용되었다. 또한, 학년에 관계없이 많은 어휘 다발이 공통적으로 사용되었다. 학력평가에서는 탈어휘 동사+명사는 다른 듣기 시험보다 현저히 적게 나타나 듣기평가에서 탈어휘 동사가 균형적으로 사용되지 않았다. Hsu(2024)는 초등학교부터 고등학교까지 영어 듣기 활동이 학습자들의 듣기 능력 향상에 미치는 효과를 분석하였다. 연구 결과, 듣기는 언어 습득의 핵심적 역할을 하며, 학생들의 말하기, 읽기, 쓰기 능력까지 향상시키는데 중요한 기반이

되었음을 확인하였다. 또한, 다양한 듣기 자료, 교실에서의 상호작용 및 평가와 피드백은 학생들의 듣기 능력 향상에 긍정적인 영향을 미치는 것으로 나타났다. 이를 통해 듣기 중심의 교수법은 학습자들의 전반적인 언어 능력 향상에 효과적이며 교육 현장에서도 활용될 필요성이 있음을 시사한다.

선행연구들을 종합해보면, 국내외 어휘와 어휘다발 연구는 주로 교과서, EBS 교재 등 읽기 영역을 중심으로 활발하게 이루어져 왔다. 이러한 연구들은 어휘다발이 학습자의 언어 능력 향상에 중요한 기여를 하며 영어 학습자료에 어휘다발을 반영할 필요가 있음을 밝혔으나, 듣기영역에 관한 어휘다발 연구는 상대적으로 부족하다. 이에 본 연구에서는 영어 듣기 평가를 중심으로 고등학교 영어 평가와 원어민 참조 코퍼스에 나타나는 어휘다발의 특성과 사용 양상을 분석하고자 한다. 본 연구는 수능, 모의평가, 학력평가 듣기 영역과 원어민 코퍼스를 분석 대상으로 삼고 어휘다발의 특성과 구조적, 기능적 특징을 비교 분석하고자 다음과 같은 연구 질문을 설정하였다.

1. 영어 듣기 평가 영역에 나타난 어휘다발의 특성은 무엇인가?
2. 영어 듣기 평가 영역에 나타난 어휘다발의 구조적 및 기능적 특성은 무엇인가?
3. 영어 듣기 평가 영역과 원어민 코퍼스에 나타난 어휘다발 간에 어떠한 유사점과 차이점이 있는가?

## 연구 방법 및 절차

### 연구 대상 코퍼스

본 연구는 대학수학능력시험, 9월 모의평가, 6월과 9월 학력평가의 듣기영역 지문을 분석 대상으로 선정하였다. 한국교육과정 평가원에서 출제한 2013년도부터 2024년도까지의 수능 및 모의평가, 학력평가 듣기 텍스트 총 75개를 코퍼스로 구축하였다.<sup>1</sup> 본 연구에서는 영어 듣기 평가(English Listening Assessment: ELA)의 어휘다발 분석을 위하여 American National Corpus(ANC)의 구어 코퍼스 중 영어듣기평가 코퍼스와 유사한 면대면 대화인 Charlotte Narrative and Conversation Collection(CNCC)를 참조 코퍼스로 사용하였다. 원어민 코퍼스인 ANC는 일상 대화체의 구조와 내용으로 이루어져, 대화체 시험 문제로 구성된 수능 듣기 영역 시험 문제 형식과도 유사하다. ANC는 구어체와 문어체의 미국 영어를 모두 포함하여 다양한 언어 사용역(일상 대화, 연설, 소설, 학술 산문 등)으로 이루어져 많은 교육 자료를 제공하며 코퍼스 기반 어휘와 담화 연구에서도 활발하게 사용되었다(Dang & Webb, 2020; Liu, 2024; Reppen, 2012).

### 분석 도구 및 절차

본 연구에서는 코퍼스 분석 프로그램인 AntConc 3.5.9을 사용하였으며, Wordlist, N-grams, Concordance, File View 기능을 사용하여 코퍼스에 나타난 어휘다발의 빈도와 사용 범위를 분석하였다. 또한, Collocate 2.1<sup>2</sup>을 사용하여 추출된 4단어 어휘다발의 상호정보(Mutual Information [MI]) 지수<sup>3</sup>를 분석하였다. 본 연구의 분석 절차는 다음과 같다. 첫째, 영어듣기평가 코퍼스와 원어민 코퍼스를 txt 파일로 구축하였다. 원어민 참조 코퍼스인 ANC 코퍼스는 웹 사이트에서 xml 확장자 형식의 파일을 사용하였다. 둘째, AntConc의 Wordlist 기능을 이용하여 영어듣기평가 코퍼스와 원어민 참조 코퍼스의 고빈도 어휘다발, 어휘 수(token), 어휘 유형(type), 어휘 다양성(TTR) 등 어휘다발 분석에 필요한 정보를 산출하였다. 어휘다발은 100만 단어당 20개 출현과 전체 파일 개수의 10%에 해당하는 파일 수로 산출하여 영어듣기평가 코퍼스와 ANC코퍼스에서 추출되는 어휘다발 출현 빈도를 각각 3회와 4회, 그리고 최소출현 파일 개수를 각각 8개와 9개로 정하였다(Chen & Baker, 2010). 셋째, 설정된 최소빈도수 기준으로 AntConc 프로그램의 N-grams 기능을 이용하여 4단어 어휘다발을 추출하였다. 넷째, Collocate 2.1을 사용하여 추출된 4단어 어휘다발의 MI 지수를 분석하였다. 마지막으로 영어듣기 평가 코퍼스와 원어민 참조 코퍼스에서 나타난

<sup>1</sup> 최근 10년치 데이터를 사용하는 이유는 현재의 트렌드를 최대한 반영하여 시의성 있는 연구와 타당성 있는 결과를 도출하기 위함이다.

<sup>2</sup> Collocate 2.1은 코퍼스에서 어휘나 언어 및 어휘다발을 빈도 또는 통계 정보(로그 우도비, 상호 정보량, t-점수)를 제공한다.

<sup>3</sup> 상호정보 지수는 단어 간 연관성을 측정하는 통계적 지표이며 MI 지수가 높을수록 단어들 사이의 연관성이 더 높으며, 낮은 지수는 우연에 의한 것일 가능성이 더 높은 것으로 볼 수 있다. 따라서 상호정보 지수는 대규모 텍스트 데이터의 언어 패턴을 분석하고 어휘 관계를 파악하는 데 적합하다(Han, 2019; Oakes, 1998; Simpson-Vlach & Ellis, 2010).

어휘다발의 구조적 유형과 기능적 유형을 분석하였다.

어휘다발의 구조적 분류는 Cortes(2004), Kim과 Lee(2019)의 분류 기준을 참고하여 범주화 하였고, 기능적 분류는 Biber 외 2인(2004)의 분류법을 참고하여 분석하였다. Cortes(2004)와 Biber 외 2인(2004)의 분류법은 어휘다발의 구조적, 기능적 특성을 체계적으로 분석하여 어휘다발 연구에서 널리 사용되고 있다. 이러한 분석을 통해, 실제 담화에서의 문법적 형태와 기능적인 역할을 명확하게 나타낼 수 있으며 어휘다발이 듣기 과정에서 자동화 처리와 의미 추론에 미치는 영향도 파악할 수 있다. 어휘다발의 구조적 분류는 다음 표 1과 같다.

**TABLE 1**  
*Structural Analysis of Lexical Bundles (Cortes, 2004)*

Categories	Examples	
Noun Phrase	Noun phrases	<i>a long time ago</i>
	Noun phrases w/ “of”	<i>one of the most</i>
	Noun phrase with post-nominal clause fragment	<i>the fact that</i>
Verb Phrase	Verb + complement	<i>was one of the</i>
Prepositional Phrase	Prepositional phrase with embedded “of” phrase	<i>as a result of</i>
	Other prepositional fragments	<i>in order to</i>
Clause		<i>it has been</i>

구조적 분류를 위하여 어휘다발을 크게 명사구, 동사구, 전치사구, 절로 분류하였다. 우선 명사구는 명사구로만 이루어진 경우, 명사구+‘of’를 포함한 경우(Noun phrases with ‘of’ phrase fragment), 명사구+‘of’ 외 표현(Noun phrase with post-nominal clause fragment)으로 구분하였다. 어휘다발이 주어 없이 일반동사 또는 be 동사만으로 이루어진 어휘다발은 동사구로 정하였다. 첫 단어가 전치사인 경우 전치사구 유형으로 분류하여 전치사구는 전치사구+명사(구)의 구조로 이루어진 어휘다발, 전치사구+‘of’를 포함한 표현 (Prepositional phrase with embedded “of” phrase), 전치사구+‘of’ 외 표현(Other prepositional fragments)으로 구분하였다. 마지막으로 주어와 동사로 이루어진 표현은 절로 구분하였다.

본 연구에서 사용된 어휘 다발의 기능적 분류는 양태 표현, 담화 조직, 지칭 표현, 특수 담화 기능이다(표 2참고). Biber 외 2인(2004)에 따르면 양태 표현(Stance expressions)은 화자의 태도나 입장에 대해 표현하는 기능으로 인식적 양태(Epistemic stance)와 태도적 양태(Attitudinal stance) 기능으로 나누어진다. 인식적 양태 기능은 *I don't know if*와 같이 주제에 대한 화자의 지식을 언급하는 기능이다. 태도적 양태 기능은 화자가 어떤 행위나 사건에 대한 태도를 나타내는 기능으로 *I don't want you* 와 같은 표현으로 하위 범주로는 바람(Desire), 의무/지시(Obligation/directive), 의도/예측(Intention/prediction), 능력/가능성(Ability/possibility)가 있다. 의무/지시 표현은 *I want you to*와 같이 타인에게 행동을 요구하거나 명령을 내리는 데 사용되며, 의도/예측 표현은 *I am going to*와 같이 의도를 전달하거나 예측할 때 나타낸다. 능력/가능성 표현은 *to be able to*와 같은 능력과 가능성을 표현한다. 바람은 *if you want to*와 같이 바람을 표현하는데 사용된다.

**TABLE 2**  
*Functional Analysis of Lexical Bundles (Biber et al., 2004)*

Categories	Sub-categories	Examples
1. Stance expressions	A) Epistemic stance	<i>I think it was</i>
	B) Attitudinal stance	
	Desire	<i>If you want to</i>
	Obligation/directive	<i>I want you to</i>
	Ability/possibility	<i>Can be used to</i>
	Intention/prediction	<i>Going to be the</i>
2. Discourse organizers	A) Topic introduction/focus	<i>Let me tell you</i>
	B) Topic elaboration	<i>You know it is</i>

Categories	Sub-categories	Examples
3. Referential expressions	A) Identification/focus	<i>One of the most</i>
	B) Specification of attributes	
	Quantification	<i>There are so many</i>
	Intangible framing attributes	<i>In the case of</i>
	Tangible framing attributes	<i>The part of the</i>
4. Special conversational functions	C) Time/place/text reference	<i>In front of the</i>
	D) Imprecision	<i>Or anything like that</i>
	A) Politeness	<i>Thank you for your</i>
	B) Simple inquiry	<i>What are you doing</i>
	C) Reporting	<i>I said to him/her</i>
	D) Emotion	<i>I am glad to</i>
	E) Feedback	<i>I agree with you</i>

담화 조직 기능(Discourse organizers)은 새로운 주제를 제시하고 주의를 유도하는 주제 소개/주의 집중(Topic introduction/focus)기능과 주제에 대한 부연설명을 하는 주제 상세화(Topic elaboration)로 나뉠 수 있다. 주제 소개/주의 집중 기능은 *What do you think*와 주제 상세화 기능은 *you know I mean*으로 분류한다. 지칭 표현(Referential expressions)은 식별/주목(Identification/focus)기능, 속성 명시(Specification of attributes)기능, 수량화(Quantification), 무형의 구성(Intangible framing attributes), 유형의 구성(Tangible framing attributes)과 시간/장소/텍스트 지칭(Time/place/text reference) 기능이 있다. 무형의 구성은 추상적인 속성을 구분하며 유형의 구성은 물리적으로 인지할 수 있는 특성을 나타낸다. 마지막으로, 특수 담화 기능(Special conversational functions)에는 공손함(Politeness), 간단한 질문(Simple inquiry), 보고(Reporting), 감정표현(Emotion), 피드백(Feedback) 등의 하위 기능이 있다.

## 연구 결과 및 논의

### 어휘다발의 특성

목표 코퍼스를 대상으로 어휘의 특성을 분석한 결과, 영어듣기평가(English Listening Assessment: ELA) 코퍼스는 총 146,469개의 토큰과 6,601개의 타입으로 구성되어 있다. 반면, 원어민 참조 코퍼스(ANC)는 199,867개의 토큰과 8,407개의 타입을 포함하고 있다. 영어듣기평가와 ANC의 타입/토큰 비율(Type-Token Ratio)은 각각 4.51, 4.21로 나타나 영어듣기평가 코퍼스가 원어민 코퍼스보다 더 다양한 어휘를 사용하고 있음을 시사하며 코퍼스 간 어휘 사용의 차이를 나타낸다. 다음 표 3은 코퍼스에 나타난 어휘다발의 상호정보를 나타낸다.

**TABLE 3**  
*Mutual Information of the Corpora*

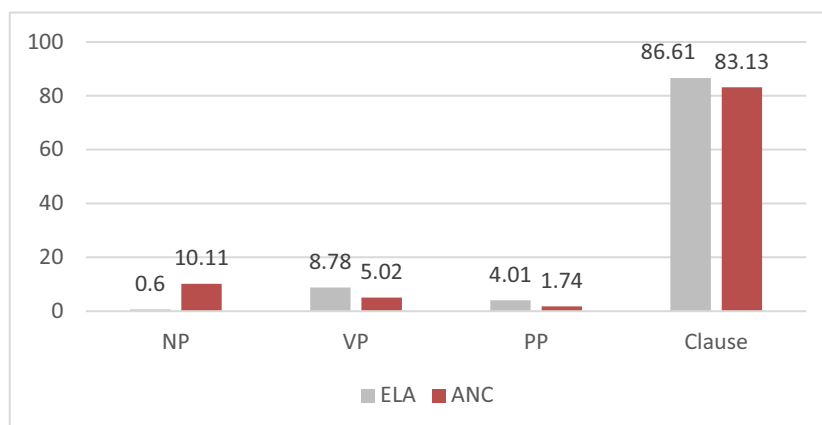
	ELA	Frequency	MI	ANC	Frequency	MI
1	<i>pay by credit card</i>	10	27.35	<i>thank you very much</i>	12	21.71
2	<i>as soon as possible</i>	8	26.74	<i>all of a sudden</i>	25	18.78
3	<i>most likely say to</i>	25	26.51	<i>do you re-member any</i>	16	17.88
4	<i>get ten percent off</i>	12	26.50	<i>a lot of fun</i>	12	17.82
5	<i>looking forward to seeing</i>	9	25.92	<i>a little bit of</i>	13	17.49
6	<i>more information please visit</i>	10	25.85	<i>or anything like that</i>	17	16.73
7	<i>a ten percent discount</i>	18	24.38	<i>to be able to</i>	13	16.70

	ELA	Frequency	MI	ANC	Frequency	MI
8	<i>have two options left</i>	10	23.94	<i>or something like that</i>	18	16.42
9	<i>will not be able</i>	15	23.88	<i>at the end of</i>	13	16.08
10	<i>this situation what would</i>	26	23.06	<i>I don't know if</i>	29	15.96

분석 결과, 영어듣기평가에서 나타난 어휘다발의 MI 지수 범위는 20.89에서 27.35로 나타났다. 영어듣기평가에서 가장 높은 MI지수를 나타낸 어휘다발은 *pay by credit card*와 *as soon as possible*이었다. 반면, ANC에서는 *thank you very much*와 *all of a sudden*이었다. 또한 ANC의 MI 지수 범위는 14.05에서 21.71로 이는 ANC에 나타난 어휘다발이 영어듣기평가에 비해 전반적으로 더 낮은 결속력을 지니고 있음을 알 수 있다. 이러한 차이는 사용역에 따른 주제와 언어 사용에 기인한 것으로 볼 수 있으며, 영어듣기평가에서는 특정한 맥락에 따라 고정된 표현을 사용하기 때문에 이러한 결과가 나타난 것으로 설명될 수 있다.

### 어휘다발의 구조적 분석

어휘 다발의 구조적 분석 결과는 다음 그림 1에 제시되어 있다.



**FIGURE 1**  
*Structural Distribution of Lexical Bundles*

그림 1에 나타난 바와 같이 영어듣기평가와 ANC에서 절이 각각 86.61%, 83.13%로 가장 많은 비율을 차지하였으며, ELA에서 동사구 8.78%, 전치사구 4.01%, 마지막으로 명사구 0.60%로 거의 나타나지 않았다.

ANC 코퍼스에서는 절 다음으로 명사구 10.11%, 동사구 5.02%, 마지막으로 전치사구 1.74%로 낮은 비율을 보였다. 이러한 결과는 대화에서 나타나는 어휘다발은 대부분이 절문이며, 약 80%의 어휘다발이 동사구의 일부를 포함하기 때문이다(Conrad & Biber, 2005).

다음 표 4는 구조적 어휘 다발 분포에 대한 카이제곱검정 결과를 나타내고 있다. 표 4에 나타난 바와 같이 영어듣기평가와 ANC 간 어휘다발의 구조적 사용 빈도 차이를 분석하기 위해 카이제곱검정한 결과, 유형에 따라 두 코퍼스 간 사용 비율이 유의미한 차이를 보였다( $\chi^2 = [275.14]$ ,  $df = 3$ ,  $p < .001$ ).

**TABLE 4**  
Chi-square Results for the Structural Distribution of Lexical Bundles

		NP	VP	PP	Clause
ELA	Observed	18	263	120	2594
	Expected	112.8	225.5	97.4	2559.3
	Residuals	<b>-8.9</b>	<b>2.5</b>	<b>2.3</b>	.7
ANC	Observed	151	75	26	1242
	Expected	56.2	112.5	48.6	1276.7
	Residuals	<b>12.6</b>	<b>-3.5</b>	<b>-3.2</b>	-1.0
Pearson Chi-square		Value	<i>df</i>	<i>Asymp. Sig. (2-sided)</i>	
		275.14	3	<.001	

영어듣기평가와 ANC에서는 명사구 표준화 잔차(residuals)가 각각(-8.9, 12.6), 동사구 표준화 잔차가 각각(2.5, -3.5), 그리고 전치사구 표준화 잔차가 각각(2.3, -3.2)로 나타났다. 이 결과에 의하면 영어듣기평가와 ANC에서 사용된 어휘다발의 구조적 유형은 통계적으로 유의미한 차이를 보였다. 이러한 유의미한 차이는 영어듣기평가에서는 명사구 사용이 ANC보다 관찰빈도(Observed)가 기대빈도(Expected)보다 더 적게, 동사구 사용은 ANC보다 더 많게, 전치사구도 ANC보다 더 많게 사용되었다고 볼 수 있다. 그리고 절은 표준화 잔차가 각각(0.7, -1.0)으로 영어듣기평가와 ANC 사이의 구조적 어휘다발 사용의 분포에 있어서 유의미한 차이가 나타나지 않았다. 이를 통해 특정 유형의 사용이 코퍼스에 따라 차이가 있음을 관찰할 수 있었다.

표 5는 어휘다발의 구조적 분석 결과이다. 명사구는 ANC코퍼스에서 더 빈번하게 사용되며 영어듣기평가에는 0.60%로 현저히 낮은 비율을 차지하였다. 두 코퍼스 모두 공통적으로 명사구 뒤에 *of* 외 다른 표현의 어휘다발이 사용되었다. 영어듣기평가 코퍼스에서는 # *percent off the total* 과 같은 한 가지 유형의 명사구만 추출된 반면, ANC에서는 *a lot of the, the rest of the* 등 총 10개의 다양한 명사구 유형이 사용되었다. 특히, 영어듣기평가 코퍼스에서 추출된 # *percent off the total*과 같은 명사구는 거의 매년 모든 시험에 사용되어 동일한 어휘다발이 빈번하게 사용되었음을 알 수 있고 ANC에서는 금액 내용의 어휘다발이 사용되지 않아 이러한 표현은 특정한 상황에서만 사용된다는 것을 시사한다. 영어듣기평가 코퍼스에서는 *to spend more than, get a #percent discount* 등 금액과 관련된 표현이 빈번하게 나타났는데 주로 직원이 고객에게 예상하는 금액대가 있는지 등의 질문에 대한 대답이나 일행과 상품을 고려하는 상황에서 의견/의도를 표현할 때 빈번하게 나타났다. 이는 듣기평가의 특성상 계산 관련 문제 유형이 자주 나타남을 반영한다.

**TABLE 5**  
Structural Analysis of Lexical Bundles

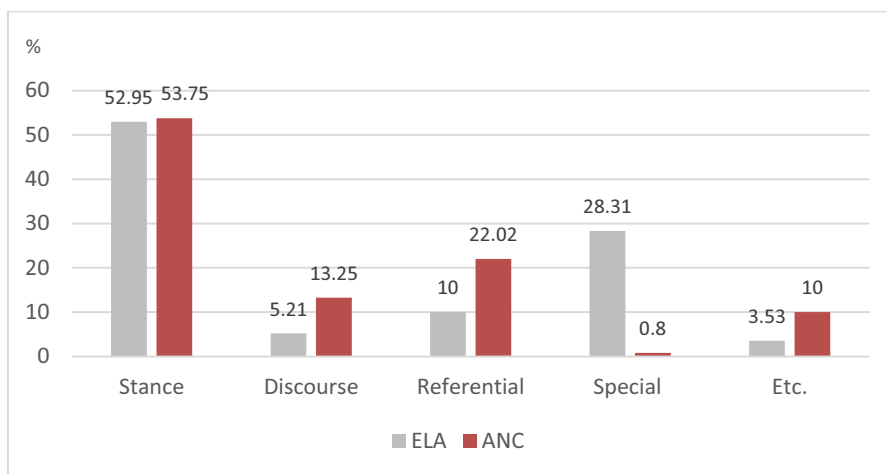
Categories	Sub-Categories	Examples	ELA (Type/Token)	ANC (Type/Token)
Noun Phrases	Noun phrases	<i>a long time ago</i>	0/0	3/55
	Noun phrases w/ "of"	<i>one of the most</i>	0/0	7/96
	Noun phrase with post-nominal clause fragment	<i>the fact that</i>	1/18	0/0
Sub-total			1/18	10/151
Verb Phrases	Verb + complements	<i>was one of the</i>	20/263	4/75
Prepositional Phrases	Prepositional phrase with embedded "of" phrase	<i>by the end of</i>	2/21	2/26
	Other prepositional fragments	<i>for a long time</i>	8/99	0/0
Sub-total			10/120	2/26
Clause		<i>I don't know</i>	153/2,594	56/1,242
Total			184/2995	72/1,494

전치사구는 두 코퍼스에서 상대적으로 낮은 비율을 보였다. 분석 결과, 영어듣기 코퍼스에서는 *of*가 포함되지 않는 구조의 어휘다발이 주로 나타났다. 이러한 어휘다발은 구체적인 의미를 전달하며 문맥에 따라 시간, 장소, 방향 등을 나타내는 데 사용되어 기능적으로는 지칭적 기능을 함유하고 있는 어휘다발에도 포함된다. 이런 구조와 기능을 함께 공유하고 있는 어휘다발의 예로서는 영어듣기평가에서는 *(in) + this situation, from am to pm, in front of the*, 그리고 ANC에서는 *at the end of the, in the middle of*가 빈번하게 나타났다. Fuster-Marquez(2022)에 의하면 학술적 산문에서는 명사구와 결합된 어휘다발(예: *the nature of*)과 전치사구와 결합한 어휘다발(예: *on the other hand*)이 많은 비율을 차지하는 반면 대화문에서는 주어 뒤에 동사구가 따라오는 형태의 어휘다발이 빈번하게 사용된다. 대화문과 안내사항, 공지 등의 내용으로 이루어진 듣기평가에서는 명사구와 전치사구가 미비하게 등장하였다는 본 연구 결과와 일관성이 있음을 알 수 있다.

두 코퍼스에서 모든 유형 중 절이 가장 빈번하게 나타났다. 절에는 주로 *I, you, we*와 같은 인칭대명사와 자주 결합하여 *be going to, didn't know* 등과 같은 동사가 나타나며 화자의 입장과 생각을 표현하는 어휘다발로 사용되었다. 두 코퍼스 모두 1인칭 대명사와 동사구가 결합한 유형이 빈번하게 나타났다. 구어체의 대화에서 나타나는 어휘다발은 대부분 동사구를 포함하며 그 구조로는 ‘인칭대명사+동사구’ 형태가 빈번하게 사용된다 (Biber et al., 2004; Hyland, 2012; Hyland & Jiang, 2018). 이는 본 연구 결과와도 일치함을 나타낸다. 이러한 이유로 영어듣기평가와 ANC 코퍼스에는 절로 분류되는 어휘다발이 다른 구조의 어휘다발과 비교하여 상당히 많은 비율을 차지하였다. 또한, 담화와 학술 산문의 차이점으로, 대화는 대부분 절의 구조로 이루어지며 학술 산문에서는 구(phrase)가 빈도 높게 나타났다는 Biber 외 4인(1999)의 연구 결과와 일치한다.

### 어휘다발의 기능적 분석

어휘 다발의 기능적 분석 결과는 다음 그림 2에 제시되어 있다. FIGURE 2에 나타난 바와 같이 고빈도 어휘다발의 기능적 분석 결과, 영어듣기평가는 양태 표현 52.95%, 특수 담화 표현 28.31%, 지칭 표현 10.00%, 담화 조직 표현 5.21%, 기타 표현 3.53% 순으로 분포 비율을 차지하였다.



**FIGURE 2**  
*Functional Distribution of Lexical Bundles*

반면, ANC에서는 양태 표현 기능이 53.75%로 수능보다 약간 높은 비율을 차지하였음을 알 수 있고 그 다음으로는 지칭 표현 22.02%, 담화 조직 표현 13.25%, 기타 표현 10%, 특수 담화 표현은 0.80%로 원어문 코퍼스에서는 특수 담화 표현이 가장 낮은 비율을 보였다. 특히 두 코퍼스에서 양태 표현이 다른 기능에 비해 상대적으로 높은 비율을 나타냈다. 이러한 이유는 두 코퍼스 모두 구어체와 대화문으로 이루어져 있기 때문에 양태 표현 기능이 높은 비율로 나타난 것으로 판단할 수 있다.

다음 표 6은 영어듣기평가와 ANC 간 어휘다발의 기능적 사용 분포 차이를 분석하기 위해 카이제곱검정한 결과를 나타내고 있다.

**TABLE 6**  
Chi-square Results for the Functional Distribution of Lexical Bundles

		Stance	Discourse	Referential	Special functions	Etc.
ELA	Observed	1584	158	303	858	107
	Expected	1595.2	237.9	422.4	581.4	173.1
	Residuals	-0.3	-5.2	-5.8	11.5	-5.0
ANC	Observed	803	198	329	12	152
	Expected	791.8	118.1	209.6	288.6	85.9
	Residuals	0.4	7.4	8.2	-16.3	71.
Pearson Chi-square		Value		df	Asymp. Sig. (2-sided)	
		655.58		4	<.001	

표 6에 나타난 바와 같이 두 코퍼스 간 사용 비율이 유의미한 차이를 보였다( $\chi^2 = [655.58]$ ,  $df = 4$ ,  $p < .001$ ). 영어듣기평가와 ANC에서는 담화조직(Discourse organizer) 표현 표준화 잔차(Residuals)가 각각 -5.2, 7.4, 지칭표현(Referential expressions) 표준화 잔차가 각각 -5.8, 8.2, 그리고 기타 표준화 잔차가 각각 -5.0, 7.1로 나타나 영어 듣기 평가에서 담화 조직표현, 지칭표현, 기타 표현 사용이 ANC보다 관찰빈도가 기대빈도보다 적게 관찰되었고, ANC에서는 이와 반대로 특수 담화 표현(Special conversational functions)의 표준화 잔차가 11.5, -16.3으로 나타나 어휘다발이 ANC보다 더 많이 나타났다. 양태 표현(Stance expressions) 표준화 잔차가 각각 -0.3, 0.4으로 나타나 영어듣기평가와 ANC의 기능적 어휘다발 사용의 분포 사이에 유의미한 차이가 나타나지 않았다.

다음 표 7은 어휘다발의 기능적 분석 결과를 나타내고 있다.

**TABLE 7**  
Functional Analysis of Lexical Bundles

Categories	Sub-categories	Examples	ELA (Type/Token)	ANC (Type/Token)
1. Stance expressions	A) Epistemic stance	<i>I think it was</i>	28/446	19/504
	B) Attitudinal stance			
	Desire	<i>If you want to</i>	20/451	2/36
	Obligation/directive	<i>I want you to</i>	10/171	0/0
	Ability/possibility	<i>Can be used to</i>	12/165	3/43
	Intention/prediction	<i>Going to be the</i>	25/351	5/220
Sub-total			95/1,584	29/803
2. Discourse organizers	A) Topic introduction/focus	<i>Let me tell you</i>	10/125	3/61
	B) Topic elaboration	<i>You know it is</i>	2/33	8/137
	Sub-total		12/158	11/198
3. Referential expressions	A) Identification/focus	<i>One of the most</i>	4/41	9/117
	B) Specification of attributes			
	Quantification	<i>There are so many</i>	6/97	5/84
	Intangible framing attributes	<i>In the case of</i>	1/10	0/0
	Tangible framing attributes	<i>The part of the</i>	0/0	1/13
	C) Time/place/text reference	<i>In front of the</i>	11/155	3/32
	D) Imprecision	<i>And stuff like that</i>	0/0	4/83
Sub-total			22/303	22/329

Categories	Sub-categories	Examples	ELA (Type/Token)	ANC (Type/Token)
4. Special conversational functions	A) Politeness	<i>Thank you very much</i>	5/56	1/12
	B) Simple inquiry	<i>What are you doing</i>	30/660	0/0
	C) Reporting	<i>I said to him/her</i>	0/0	0/0
	D) Emotion	<i>I am glad you</i>	7/83	0/0
	E) Feedback	<i>is a good idea</i>	3/59	0/0
Sub-total			45/858	1/12
5. Etc.		<i>to go to the</i>	11/107	10/152

표 7에 제시된 바와 같이 양태(Stance) 표현에서는 영어듣기평가(ELA) 코퍼스는 인식적(Epistemic), 바람(Desire), 의도/예측(Intention/prediction), 의무/지시(Obligation/directive), 능력/가능성(Ability/possibility) 순으로 그리고 원어민 코퍼스는 인식적, 의도/예측, 능력/가능성, 바람, 의무/지시 순으로 빈번하게 나타났다.

두 코퍼스에서 공통적으로 인식적 기능이 가장 높은 비율을 차지하였는데, 영어듣기 평가와 ANC 양태 기능에서 인식을 나타내는 *I don't, didn't know, I think it is* 등을 포함한 어휘다발이 가장 빈번하게 사용되었다. *Don't*와 결합해서 자주 사용되는 동사는 *think*와 *know*이므로 화자의 인지 상태와 입장이 빈번하게 표현되었음을 알 수 있다. 이러한 특성은 고등학교 교과서 및 원어민 코퍼스를 분석한 Yang(2020)의 결과와 일치한다. Biber 외 4인(2004)에 따르면 양태표현 기능의 어휘다발은 절로 이루어지는 경향을 보인다고 하였는데, 본 연구에서도 인식적 기능의 어휘다발은 모두 절로 구성되었음을 알 수 있다. 두 코퍼스의 차이점으로는 ANC에서는 영어듣기평가보다 과거 시제 *didn't*와 포함된 어휘다발이 빈번하게 나타나며 ANC에서 더 다양한 유형으로 사용되었음을 알 수 있다. 또한, ANC에서는 *um, I don't, uh I don't, I don't really*는 구어체의 특징으로 빈번하게 나타났으며 담화표지어 *um, uh, like*와 같은 표현이 대부분 다른 단어나 구(phrase)와 함께 결합하여 빈번하게 사용되었다. 이러한 표현은 주로 문장이나 절의 앞 부분에 나타나며, 불확실성을 표현하거나 화자가 다음에 말할 내용을 정리하는 시간을 준다. 이는 대화의 구조로 구성된 영어듣기평가에서 화자와 상대방이 서로의 입장을 표현하여 상호작용하는 대화의 특성이라고 볼 수 있다.

바람 기능은 ANC 보다 영어듣기평가에서 현저하게 높은 비율로 사용되었으며 영어듣기평가에서 바람과 관련된 동사가 많이 출현한 것으로 나타났다. 두 코퍼스 모두에서 *don't want to* 표현이 빈번하게 나타나며 이는 화자의 의견을 표현하기 위해 구어체에서 자주 사용되어 구어적 특성을 반영한다고 볼 수 있다. 영어듣기 평가에서는 *I don't want to*가 동사 *spend, pay, be*와 많이 사용되었으며 이는 물건을 구매하는 상황에서 의사를 표현하는 대화 내용으로 구성되어 있는 문체 유형의 특성 때문에 이러한 어휘다발이 빈번하게 나타났음을 알 수 있다. 영어듣기평가에서 가장 빈번하게 나타난 바람의 기능과 관련된 어휘다발은 *I'd like to*와 *I am looking for*이며 구매하고자 하는 물건을 찾을 때나 정보를 요청할 때 사용되는 표현으로 사용 빈도가 높게 나왔다. 위의 예시와 같이 *I'd like to*는 시험에 매년 출제될 정도로 빈번하게 나타났다. 반면, 원어민 코퍼스에서 바람 기능의 사용은 바람을 표현하는 직접적인 단어 *want*로만 반복적으로 나타났다. 이처럼 영어듣기평가에서 기본적으로 사용되는 *want* 외에 *looking forward, wish, can't wait*와 같이 비교적 다양한 표현으로 사용되었음을 알 수 있다.

영어듣기평가는 능력/가능성 기능이 다른 범주 중 가장 낮은 비율로 나타났다. 영어듣기평가 코퍼스와 ANC에서 공통으로 *be able to*와 *I can't*를 포함한 어휘다발이 나타났다. 영어듣기평가에는 *be able to* 앞에 *will not, won't*이 항상 사용되어 부정의 의미로 나타난 것을 알 수 있었다. 반면, 원어민 코퍼스에서 *to be able to*는 주로 앞에 *am going to/would love/want/used/had* 등과 결합하여 이후에는 동사 *go, have, use, tell, decide*와 나타나 다양한 맥락에서 긍정의 의미로 사용되었음을 알 수 있다. 학술 담화 코퍼스에서도 가장 빈번하게 나타나는 어휘다발 중 하나인 *be able to*는 그 의미의 변화가 거의 없이도 생략될 수 있으나, 이러한 어휘다발을 사용함으로써 뒤따라 나오는 구문을 유연하게 표현해주는 역할을 하기도 한다(Kim & Lee, 2019; Yang, 2020).

의도/예측 표현 유형으로는 두 코퍼스 공통으로 *be going to*와 *am trying to*를 포함한 어휘다발이 빈도 높게 사용되었으며, 미래의 행동이나 계획을 표현하는 의미를 가지고 있다. 따라서, 의도/예측에 나타난 어휘다발은 실제 원어민이 사용하는 어휘다발과 유사하다고 볼 수 있다. 듣기평가에서는 *I am going to, I was going to*와 같이 1인칭 대명사와 *going to*가 나타났다. 또한, 질문하기 위해 주로 2인칭 대명사 *you*를 사용하여 *are you going to*와 같은 어휘다발도 함께 사용되었다. 이는 두 명의 대화 상대가 대화를 나누는 상황이 평가를 위한 상황 설정에 필수적이기 때문이다. 반면, 평가의 상황이 고려되지 않는 일반 대화문으로 구성된 ANC에서는 수능에서 사용되는

제한된 형식의 질문 형태보다는 1인칭 대명사뿐만 아니라 *we, you, she, they* 등과 같은 다양한 형태의 대명사와 결합되며 문법상으로도 과거 시제가 사용되는 다양한 구조와 형식의 어휘다발이 사용되었다는 것을 확인할 수 있었다.

의무/지시 표현에는 *have to, should*와 같은 단어가 빈번하게 사용되었다. *I, you, we they*와 같은 1인칭대명사가 *wait, pay, use* 등과 같은 동사와 결합하여 사용되었다. 본 연구의 양태 표현의 의무 기능은 구어와 문어 사용역에서 빈번하게 등장하며 2인칭 *you* 가 종종 포함되어 *I want you to, you have to be, you need to take, you might want to* 와 같이 청자에게 특정 행동을 지시하는 표현으로 사용된다는 연구 결과와 일치한다(Biber et al., 2004; Biber & Barbieri, 2007; Pan et al., 2016).

담화 조직 기능에서는 두 코퍼스에 주제 상세화와 주제 소개/주의집중 순으로 나타났다. 영어듣기평가에서 새로운 주제 소개/주의 집중을 표현하는 *take a look, today I'd like, let me show you, let's look at*이 높은 빈도로 사용되었다. 영어듣기평가에서는 *take a look at*이 빈번하게 출현하였는데 *take+*명사구의 구조로 빈번하게 사용되어, (목적어)를 하다는 의미로 나타났다. 영어듣기평가에서 *today I'd like*는 *tell, inform, introduce, begin* 등과 같이 결합되어 연주회, 뉴스, 대학교, 이벤트의 안내나 공지사항에서 빈번하게 나타났다. 듣기 평가의 특성상, 주어진 상황이 화자와 청자 간의 대화문과 독백문의 구조로 이루어지기 때문에 1인칭이나 2인칭 주어가 사용되었고 주제 소개와 주의 집중과 관련된 어휘 다발이 나타났음을 알 수 있다. 반면, ANC에서는 *when I was in, when I was* 와 같이 대화 주제를 소개하거나 집중을 유도하기 위한 어휘다발이 포함되며 *when*절이 주절의 주제를 이끌고 있다. *When I was a* 와 *When I was in*은 화자가 어렸을 때, 과거의 시점을 회상하며 대화를 이어나가는 상황에서 나타났다. 이처럼 *when I was*는 시간을 나타내는 부사절의 형태로 사용되어 주절의 상황을 구체적으로 설명하거나 화자의 생각이나 감정을 표현하는데 나타났다. 주제 상세화 기능으로 영어듣기평가에서는 *you're right I*가 사용되었으며 앞서 언급된 상대방의 이야기에 화자의 의견을 덧붙여 부연적으로 설명하거나 동의하는 내용을 포함한다. *You're right I*는 주로 *will book, will buy, will go, will apologize* 와 같이 미래형으로 사용하여 상대방이 한 말에 동의하며, 동의한 내용을 바탕으로 행동을 하겠다는 표현이 나타났다. 따라서, 이러한 어휘다발은 우호적인 분위기를 이끌어 내는 구문으로 대화 상대 간의 협력적인 분위기를 조성하여 의사소통을 증진시키는 역할을 한다고 볼 수 있다. 반면, 원어민 코퍼스에서 *that's what I, that's why I*는 특정한 행동이나 사실을 설명할 때와 원인을 제시할 때 사용되었다. 이러한 차이는 듣기영역의 텍스트가 읽기영역 텍스트에 비해 문장의 길이가 짧기 때문에 연결어 없이도 상황을 파악할 수 있기 때문으로 이는 중학교와 고등학교 교과서에 나타난 Yang(2020)의 연구결과로도 그 이유를 유추해 볼 수 있다. 한편, ANC에서 나타난 *you know*와 같은 표현은 Simpson-Vlach과 Ellis(2010)의 연구에서도 나타나며 화자가 발화를 하거나 수정할 때 사용하는 표현이다. 이와 같이 원어민 코퍼스에서 듣기평가 코퍼스보다 더 자연스러운 대화의 흐름을 나타나고 있다는 것을 알 수 있다.

지칭(Referential) 표현에서는 영어 듣기 평가의 하위 범주로 시간/장소/텍스트 지칭 비율이 51.16%로 가장 높았으며 수량화 32.01%, 식별/주목 13.53%, 무형의 구성이 3.30%의 비율을 차지하였다. 그러나 유형의 구성기능과 부정확성 기능을 하는 어휘다발은 영어 듣기 평가에서는 나타나지 않았다. 반면, ANC에서는 식별/주목 37.50%, 수량화 26.92%, 부정확성 21.15%, 시간/장소/텍스트 지시 기능 10.26%, 유형의 구성 4.17%의 비율을 차지하였다. 그리고 영어 듣기 평가에서 3.30%의 비율을 차지했던 무형의 구성은 ANC에서는 나타나지 않았다.

식별/주목으로 분류되는 어휘다발은 공통적으로 *is/was one of the, oh there is/was a, and it was a* 와 같이 주로 *be* 동사와 결합한다. *This is a, it is*는 주로 명사구와 결합하며 *one*과 *it*은 사람이나 특정한 대상과 개념을 지칭할 때 사용되었다. 이러한 어휘다발은 모두 특정 명사구에 초점을 두어 구문의 앞부분에 위치한다. *Is/was one of the* 유형은 두 코퍼스에서 모두 지칭 기능의 유형 절반 이상이 *one of the most, one of the best* 와 같이 최상급의 의미로 빈번하게 나타났다. 특히 *one of the*가 포함된 표현들은 구체적인 정보를 제공하며 중요한 점을 강조하였다. 이러한 어휘다발은 주로 영어듣기평가의 독백문에서 나타나 다수의 청자들에게 구체적인 정보나 새로운 사실을 제공할 때 사용된 표현으로 볼 수 있다. 원어민 코퍼스에서는 *one of the* 앞에 지시대명사 *that*이 빈번하게 등장하여 앞서 제시된 특정한 대상이나 경험을 지칭하는 역할을 한다고 볼 수 있다.

두 코퍼스 공통으로 사용된 수량화 표현은 *have a lot of*이다. Simpson-Vlach과 Ellis(2010)에 따르면 속성 명사 기능은 대부분 *a lot of* 와 같이 명사구 앞에 양을 구체화한다고 설명하였다. 두 코퍼스의 차이점을 살펴보면, 영어듣기평가에서 할인이나 금액 정보를 정확히 파악해야 하는 문제 유형에서 *get a ten percent, # percent off the total, off the total price, dollars*와 같은 어휘다발이 빈번하게 나타났음을 알 수 있다. 따라서 영어평가에서 수량화 표현의 범위를 확대하여 사용할 필요성이 있음을 제시한다. 또 다른 차이점으로는, ANC에서는 수량화를 나타내는

*a lot of* 를 포함한 어휘다발이 빈번하게 등장하며 *a little bit* 와 같이 구어체의 표현도 많이 사용되었다. 이는 학술적인 상황보다 구어체나 비격식적인 상황에서 자주 사용되었음을 보여준다(Biber & Barbieri, 2007; Choi, 2019).

무형의 구성 기능은 보이지 않는 추상적인 특성이나 속성을 지칭하며 *a big fan of*와 같은 어휘다발이 영어듣기평가에서만 등장하였다. 유형의 구성 기능은 *the rest of the* 와 같이 구체적인 특성을 나타내는 어휘다발이며 원어민 코퍼스에서만 나타났다. 유형의 구성 기능은 무형의 구성 기능에 비해 상대적으로 낮은 빈도로 사용되었다. 이는 구체적인 설명이 필요한 학술 산문과 다르게 구어 특성의 차이점으로 볼 수 있기 때문에 구어체인 영어듣기평가에서는 전체 지칭 표현의 분류에서 매우 미비한 역할을 하는 것으로 보인다.

부정확성 기능에서 *and stuff like that, or something like that*와 같은 어휘다발은 영어듣기평가에서는 나타나지 않았고 원어민 코퍼스에서만 사용되었다. 이런 어휘다발은 의미전달에 필요한 내용이 아니며 학습자들의 영어 듣기 능력을 측정하는데 중요한 요소가 아니라는 출제자들의 의도 때문이라고 판단된다. 이러한 어휘다발은 영어듣기평가에도 점차 포함시켜 학습자들의 자연스러운 대화를 이끄는 학습을 유도할 필요성이 있다고 본다.

영어듣기평가에서 지칭 기능 중 시간/장소/텍스트 지시 표현이 가장 높은 분포 비율을 보였다. *In front of the, for a long time* 와 같이 위치나 장소, 기간을 구체적으로 표현할 수 있는 어휘다발이 가장 빈번하게 나타났으며, 두 코퍼스에서는 공통적으로 다기능 표현인 *the end of*가 나타났는데 앞에 나오는 전치사 *by*와 *at*에 따라 미세하게 다른 의미로 제시되었다. 한편, ANC에서는 *at the end of* 와 *this road, street, line, day, night*와 같은 장소/위치를 나타내는 어휘와 빈번하게 사용되었다. 이러한 어휘다발은 다기능적(multi-functional) 특성을 보이며, 특정 맥락에 따라 장소, 시간, 또는 텍스트 지시를 제시하는 기능을 한다. 따라서 이러한 어휘다발은 상황에 따라 다양한 지시적 기능의 역할을 한다고 보여준다.

마지막으로 특수 담화 기능(Special conversational functions)을 살펴보면, 영어듣기평가 코퍼스에서는 간단한 질문(Simple inquiry)이 74.43%로 압도적으로 많이 나타났으며 감정(Emotion)표현, 공손함(Politeness), 피드백(Feedback) 순으로 나타났다. 반면, ANC에서는 공손함 기능만 나타났으며 두 코퍼스에서 간단한 보고(Reporting) 기능은 나타나지 않았다. 따라서 특수 담화 기능은 구어체를 사용하는 구어 코퍼스의 특성을 반영하여 ELA에 빈도 높게 사용되었음을 알 수 있다.

공손함 기능에 두 코퍼스가 공통적으로 *thank you*가 나타났으며 영어듣기평가에서는 *thank you for* 뒤에 주로 *advice, time, help, cooperation* 등이 가장 빈번하게 등장하였다. 이는 ANC에서와도 비슷한 결과를 보였다. 영어듣기평가에서 *thank you for listening* 은 주로 공지사항, 강연, 안내 상황 같은 독백문에서 끝맺음 표현으로 사용되었다.

간단한 질문은 특수 담화의 다른 하위 기능 중 가장 높은 비율로 나타났다. 이 유형의 어휘다발로는 *why don't you/we, may I help you, what are you doing* 등이 있으며 *may I help you*는 앞에 *how*와도 결합하여 주로 여행사, 매장, 호텔, 서비스 센터 등의 문맥에서 사용되었다. 이러한 어휘다발은 듣기 평가의 특성상 두 사람의 대화에서 대부분 질문과 대답의 구조로 이루어진 문제로 구성되어 있어 빈번하게 출현된 것으로 보인다. 반면 일상 생활의 대화로 구성된 ANC에서는 간단한 질문의 기능을 포함하는 어휘다발이 나타나지 않았다.

감정표현 어휘다발은 *I'm so excited, I'm glad to, I'm glad you, sorry to hear that*이 가장 빈번하게 사용되었으며 원어민 코퍼스에서는 나타나지 않았다. 피드백 기능에 포함되는 어휘다발 *that is a good/great*는 *plan*과 *idea*와 가장 빈번하게 결합하였다. 또한, 보고 기능에 속하는 어휘다발은 두 코퍼스에서 나타나지 않았다. 마지막으로, 단순 보고 기능의 어휘다발은 영어듣기평가와 ANC에서 모두 나타나지 않았다. D. Choi(2015)가 강조한 *바와 같이 It is said that, It is believed that* 등과 같은 표현은 주로 문어체에서 사용된다.

## 결론

본 연구에서는 고등학교 영어 듣기 평가에 나타난 어휘다발의 특성을 밝히고자 대학수학능력시험, 모의평가, 학력평가의 듣기 영역에서 가장 빈번하게 나타난 4단어 어휘다발을 추출하고 구조적 그리고 기능적 특성을 분석하였다. 아울러, 원어민 코퍼스인 ANC에 나타난 어휘다발의 구조적, 기능적 범주와 특성을 비교 분석하여 분포 양상을 파악하였다. 연구 결과, ANC에서는 적은 수의 어휘다발이 반복적으로 사용된 반면 영어듣기평가에서는 다양한 유형의 어휘다발이 사용되었다. 이러한 결과는 영어듣기평가는 특정 어휘나 표현을 의도적으로 포함하여 많은 수의 어휘다발이 반복적으로 사용된 것을 알 수 있다. 이는 대부분 교육과정에서 중요시하는 필수

어휘 및 어휘다발을 반영한 평가임을 알 수 있었다. 더 나아가, 평가는 실제 언어 사용을 반영하면서도 특정한 주제의 텍스트에 자주 사용되는 표현이 정해져 있기 때문에 고정된 표현이 많이 사용된 것으로 보인다. 또한, 구조적 분석 결과, 두 코퍼스에서 공통적으로 절이 가장 빈번하게 사용되었고 다음으로는 동사구, 전치사구, 명사구 순으로 나타났다. 이는 대화의 형식으로 이루어져 두 코퍼스 간 유사한 구조적 특징이 나타난 것으로 볼 수 있다. 또한, 영어듣기평가는 ANC 보다 더 많은 어휘다발이 나타났지만 이는 제한된 어휘가 반복적으로 사용된 결과라고 할 수 있다. 즉, 영어듣기 평가는 평가 유형상 특정한 어휘다발이 반복적으로 빈번하게 사용되었음을 의미한다. 따라서 평가의 역류 효과를 고려한다면 학습자들의 의사소통 능력을 향상시키기 위해서는 영어듣기 평가에 어휘다발을 다양하고 균형있게 제시하여 풍부한 어휘 표현을 학습할 수 있도록 할 필요가 있다. 기능적 분석 결과에 따르면, 두 코퍼스에서 모두 양태 표현이 가장 높은 빈도로 나타났고 듣기평가에서는 특수 담화, 지칭 표현 순으로 사용되었다. 반면, ANC에서는 지칭 표현, 담화 조직 표현 순으로 나타났다. 이는 구어, 교실 수업, 일상 대화를 분석한 Biber와 Barbieri(2007)와 Csomay(2013)의 결과와 일치한다. 또한, 영어듣기 평가와 ANC는 주로 서로의 의견을 전달하는 구어체와 대화문으로 이루어져 있기 때문에 이러한 결과가 나타났음을 알 수 있다. 따라서, 영어듣기평가 코퍼스와 ANC 간의 차이점이 하위 기능들의 분포를 통해 드러났다고 할 수 있었다. 영어듣기평가에서 ANC보다 특수 담화 표현이 더 높은 비율로 나타났는데 이는 간단한 질문이 평가에서 더 많이 사용되었기 때문이라고 판단된다. 또한, 지칭 표현은 영어듣기평가 보다 ANC에서 더 높은 비율로 나타났음을 확인할 수 있었다. 이러한 결과는 하위 기능의 부정확성과 식별/주목 출현의 차이로 나타났기 때문이라고 볼 수 있다.

영어듣기평가와 ANC의 어휘다발 유형 및 유사점과 차이점을 살펴보았을 때, 영어듣기평가에서 어휘 다발이 빈번하게 나타났으며, 특정 범주 기능을 과도하게 사용하는 경향이 나타났다. 영어듣기평가에서 구조적으로는 절과 기능적으로는 양태와 특수 담화 기능이 다른 범주에 비해 많이 사용되어 어휘다발의 균형적 사용이 나타나지 않았다. 이는 대화 형식 중심으로 구성된 영어듣기평가의 문제 유형과 구어적 특성을 반영한다고 볼 수 있다. ANC의 경우에는 영어듣기평가와 달리 다양한 담화표지어와 부정확성 기능의 어휘다발이 나타났다. 이러한 어휘다발은 대화의 자연스러운 흐름을 촉진시키며 화자의 의도를 전달할 수 있는 역할을 한다.

영어 듣기 평가는 평가 도구임과 동시에 교과서만큼 매우 중요한 입력자료이자 학습자료이다. 따라서 수능, 교재, 평가에 원어민들이 실제 사용하는 표현과 어휘다발을 다양하게 포함시켜 입시를 준비하는 학습자들이 어휘다발의 중요성을 인식하고 학습할 수 있도록 해야 함을 시사한다. 즉, 명사구, 전치사구와 지칭 표현, 담화 조직 표현 기능과 관련된 어휘다발을 대화문에 더 포함시켜 균형적으로 사용될 수 있도록 할 필요성이 있다. 또한, 본 연구의 결과는 어휘 학습에 있어서 개별 어휘 의미를 위주로 학습하는 것보다 어휘의 실제 사용에 중점을 두는 것이 매우 중요하다는 것을 시사한다. 즉, 문맥에서 어휘를 학습하고 어휘 사용 능력을 개선시킬 수 있도록 단어 및 어휘다발을 중심으로 학습하는 방안을 모색해야 된다는 것이다. 본 연구의 구조적 및 기능적 분석 결과를 통해, 교수자들은 학습 목적에 따라 어떤 유형의 어휘다발을 우선적으로 지도할지 결정하는데 참고하고 수업을 기획할 수 있으며, 다양한 매체를 활용하여 실생활에 사용되는 어휘 및 어휘다발을 학생들에게 노출시켜 어휘다발의 사용법을 지도할 수 있다. 이와 같이 어휘다발 연구 결과는 학습자, 교수자, 연구자 모두에게 향후 교육과 연구 현장에서 적극적으로 활용될 수 있으리라 기대된다.

## References

- Appel, R. (2022). Lexical bundles in L2 English academic texts: Relationships with holistic assessments of writing quality. *System*, 110, 102899.
- Biber, D. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press, 2, 230-239.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Harlow: Pearson.
- Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research*, 18(1), 54-74.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14(2), 30-49.

- Choi, Da-Yeon. (2015). *A comparative corpus analysis of lexical bundles in high school English textbooks and BNC baby* [Unpublished MA thesis]. Korea University.
- Choi, Hye-ji. (2019). *A comparative analysis of lexical bundles in Korean middle school English textbooks and listening proficiency tests* [Unpublished master's thesis]. Korea National University of Education.
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22-29.
- Conrad, S., & Biber, D. (2005). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56-71.
- Cortes, V. (2004). Lexical bundles in published and student writing in history and biology. *English for Specific Purposes*, 23(4), 397-423.
- Coxhead, A., Dang, T. N. Y., & Mukai, S. (2017). Single and multi-word unit vocabulary in university tutorials and laboratories: Evidence from corpora and textbooks. *Journal of English for Academic Purposes*, 30, 66-78.
- Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics*, 34(3), 369-388.
- Dang, T., & Webb, S. (2020). Vocabulary and good language teachers. In C. Griffiths & Z. Tajeddin (Eds.), *Lessons from good language teachers* (pp. 203-218). Cambridge University Press.
- Fuster-Márquez, M. (2022). Applying a corpus-driven approach in linguistic analysis: The case of lexical bundles and phrase frames. In S. Maci & M. Sala (Eds.), *Corpus linguistics and translation tools for digital humanities: Research methods and applications* (pp. 62-81). Bloomsbury Academic.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Clarendon.
- Han, Sumi. (2019). Frequency, MI, and congruency in collocation processing by Korean EFL learners: Evidence from reading aloud. *English Linguistics*, 19(3), 325-346.
- Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further development in the lexical approach* (pp. 47-69). Language Teaching Publications.
- Hsu, Y. H. (2024). An analysis of the effectiveness of listening activities in improving English learning/reading (L2 Language-English). In *SHS Web of Conferences* (pp. 1-10). EDP Sciences. <https://doi.org/10.10541/shsconf/202420002022>
- Hyland, K., & Jiang, K. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383-407.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- Kim, Youn Kyong., & Lee, Eun-Joo. (2019). A corpus-based comparative analysis of formulaic sequences in high school English reading assessment texts. *Foreign Language Education*, 26(2), 99-123.
- Kwon, Ye-Eun. (2021). An analysis on lexical bundles of college students' speaking scripts: Based on a self-introduction task. *Modern English Education*, 22(4), 79-89.
- Lee, Eun-Joo. (2009). A corpus-based study of the Korean EFL learners' use of formulaic sequences. *Foreign Languages Education*, 16(2), 321-340.
- Lee, Soo Hyun. (2022). *The corpus-based analysis of lexical bundles in the 2015 revised curriculum high school English textbook dialogue: Comparison with BNC baby*. [Unpublished master's thesis]. Ewha Womans University.
- Leech, G. (1992). Corpora and theories of linguistic performance. In J. Startvik (Ed.), *Directions in corpus linguistics* (pp. 105-122). Mouton de Gruyter.
- Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward*. Language Teaching.
- Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 126-154). Oxford University Press.
- Liu, C. Y., & Chen, H. J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, 58, 122-137.
- Liu, Y. (2024). Sport or politics? A corpus-based critical discourse analysis of Chinese and American media coverage of the Paris 2024 Olympic Games. *Open Journal of Social Sciences*, 12(10), 168-184.
- McCarthy, M. (1990). *Vocabulary*. Oxford University Press.
- McCarthy, M., & O'Dell, F. (2005). *English collocations in use: How words work together for fluent and natural English: Self-study and classroom use*. Cambridge University Press.
- Nation, I. S., & Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Oakes, M. P. (1998). *Statistics for corpus linguistics*. Edinburgh University Press.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Park, Myung ho., & Lee, Dong Ju. (2021). A corpus-based study on vocabulary use in English listening tests for high school students.

- Brain, Digital, & Learning*, 11(3), 447-468.
- Reppen, R. (2012). Lessons from the American National Corpus. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 204-212). Continuum.
- Shin, Dongkwang, Chon, Yuah, Lee, Shinwoong, & Park, Myongsu. (2018). A comparison of single word and multi-word unit profiles in spoken and written corpora of Korean learners and English native speakers. *Journal of the Korea English Education Society*, 17(2), 93-112.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225.
- Wang, W., & Csomay, E. (2024). Constructing proximity in popularization discourse: Evidence from lexical bundles in TED talks. *English for Specific Purposes*, 73, 95-109.
- Wu, Y. (2021). Lexical bundles in English EU parliamentary discourse-variation: Across interpreted, translated, and spoken registers. *Compilation & Translation Review*, 14(2), 37-86.
- Yang, Won-Na. (2020). *A corpus-based analysis of lexical bundles: Focusing on reading texts in English textbooks of the 2015 revised curriculum* [Unpublished master's thesis]. Korea National University of Education.