

영국 학술 구어 코퍼스(BASE): 구축, 연구 활용 및 AntConc 활용 교육

정영옥
(한국외국어대학교)
정채관
(국립인천대학교)

Jong, Y. O., & Jung, C. K. (2026). The British Academic Spoken English Corpus (BASE): Compilation, research applications, and AntConc-based pedagogy. *Asia Pacific Journal of Corpus Research*, 7(1), 7-12.

This article reexamines the British Academic Spoken English Corpus (BASE) as a resource for research on university lectures and seminars and corpus-based language education. Drawing on the BASE manual, corpus documentation, and five representative studies, it reviews BASE's compilation, disciplinary balance, transcription conventions, versions, research applications, and pedagogical potential. The official release contains 1,644,942 tokens from 160 lectures and 39 seminars recorded at two British universities between 1998 and 2005. The review shows that BASE supports research on lexical bundles, pragmatic functions of just, relevance markers, lexical coverage, and tag questions, linking vocabulary, discourse organization, and lecturer-student interaction. Its strengths include coverage of four disciplinary domains, inclusion of both lectures and seminars, and annotation of speaker and nonverbal information. However, differences among the original XML files, research-specific datasets, and platform versions affect document and token counts. Uneven annotation, limited seminar data, normalized transcription, and restriction to two institutions constrain interpretation. Three AntConc activities are proposed: comparing just across lectures and seminars, contrasting BASE with MICASE, and analyzing relevance markers. The article argues that BASE remains valuable when researchers document versions and preprocessing decisions and treat findings as evidence about particular British academic settings rather than academic speech in general.

Keywords: British Academic Spoken English Corpus (BASE), Academic Spoken English, Corpus-Based Discourse Analysis, AntConc, Data-Driven Learning

1. 서론

대학의 학술 활동은 논문과 교재 같은 문서 자료뿐 아니라 강의, 세미나, 토론 및 질의응답과 같은 구어 담화를 통해서도 이루어진다. 교수자는 강의에서 개념과 이론을 설명하고 중요 내용을 부각하며, 학생의 이해를 확인한다. 학생은 질문과 응답을 통해 지식을 재구성하고 학문 공동체의 담화 관습을 습득한다. 따라서 대학의 실제 언어 사용을 이해하려면 학술 문어와 함께 학술 구어를 체계적으로 분석할 필요가 있다. British Academic Spoken English Corpus(이하, 'BASE')는 이러한 필요에 따라 영국 대학의 실제 강의와 세미나를 녹음하고 전사하여 구축한 학술 구어 코퍼스이다. BASE가 구축되기 전에도 미국 대학의 학술 구어를 수집한 Michigan Corpus of Academic Spoken English(이하, 'MICASE')가 존재했지만, 영국 대학의 강의와 세미나를 여러 학문 분야에 걸쳐 체계적으로 비교할 수 있는 자료는 충분하지 않았다. BASE는 이러한 자료상의 공백을 보완하고 MICASE와 비교 가능한 영국 학술 구어 자료를 제공하기 위해 구축되었다(Thompson & Nesi, 2001).

BASE 구축은 Hilary Nesi와 Paul Thompson이 주도하였으며, 이들에 따르면 University of Warwick과 University of Reading에서 수집한 자료를 중심으로 이루어졌다. 녹음 자료는 1998년부터 2005년 사이에 생산되었고, 공식 데이터베이스에는 구축 시기가 2000년부터 2005년까지로 기록되어 있다. 코퍼스는 2008년에 공개되었다. 공식 원자료는 강의 160개와 세미나 39개, 총 199개의 담화 사건으로 구성되며, 전체 규모는 1,644,942토큰이다. 자료는 인문예술, 생명/의학, 물리과학, 사회과학이라는 네 개의 넓은 학문 분야를 비교적

고르게 포함한다(Corpus Resource Database, 2014; Nesi & Thompson, 2006).

BASE의 명칭에서 ‘academic’은 문장이 어렵거나 격식적이라는 뜻이 아니라 대학의 강의와 세미나라는 학술적 상황에서 담화가 생산되었다는 의미이다. ‘British’ 역시 모든 화자가 영국 국적자이거나 영국 영어 원어민임을 뜻하지 않는다. 영국의 두 대학에서 수집된 학술 담화라는 지리적, 제도적 성격을 가리킨다. 따라서 BASE에서 관찰된 언어적 특성을 영국인 전체의 언어, 모든 영국 대학의 학술 영어 또는 영어 원어민 전체의 학술 구어로 확대해서는 안 된다. 본고는 BASE의 구축 배경과 표집 설계, 자료 구성 및 전사 원칙을 검토하고, 이 코퍼스가 실제 연구에서 어떻게 활용되어 왔는지를 대표 연구를 중심으로 살펴보는 데 목적이 있다. 아울러 원자료와 가공판의 차이, 분석 결과의 일반화 범위 및 재현성 문제를 논의하고, AntConc(Anthony, 2026)를 활용한 연구와 교육 활동을 제안하고자 한다. 이를 통해 BASE가 학술 어휘를 조사하는 자료를 넘어 강의 구조, 담화 기능, 교수자와 학생의 상호작용 및 대학 영어교육을 연결하는 연구 자원임을 제안하고자 한다.

2. The British Academic Spoken English (BASE)

2.1. BASE의 구성과 표집 설계

Dang과 Webb (2014)에 의하면 BASE의 기본 분석 단위는 하나의 강의 또는 세미나로 이루어진 담화 사건이다. 공식 원자료에는 강의 160개와 세미나 39개가 포함되어 있다. 강의는 인문예술, 생명/의학, 물리과학 및 사회과학 분야에 각각 40개씩 배분되어 있다. 세미나는 물리과학 분야에 9개, 나머지 세 분야에 각각 10개가 포함되어 있다. 이에 따라 인문예술, 생명/의학 및 사회과학에는 각각 50개의 담화 사건이 있고, 물리과학에는 49개가 있다. 이러한 구성은 네 개의 넓은 학문 분야를 비교하는 데 유리하다. 특정 표현이 인문예술 강의와 물리과학 강의에서 어떻게 다르게 사용되는지, 독백적 성격이 강한 강의와 교수자 및 학생의 상호작용이 활발한 세미나에서 담화 표지가 어떻게 달라지는지를 분석할 수 있다. 그러나 BASE는 영국 영어 전체를 대표하도록 설계된 일반 균형 코퍼스가 아니다. 일상 대화, 언론, 법률 및 상업 담화를 포함하지 않으며, 학술 구어 중에서도 강의와 세미나에 집중되어 있다. 연구실 회의, 개별 지도, 학회 발표, 실험실 활동과 같은 다른 학술 담화는 충분히 반영하지 못한다. 공식 규모인 1,644,942토큰도 고정된 절대값으로 보기 어렵다. 숫자, 축약형, 하이픈 표현, 문장부호, 화자 코드 및 Extensible Markup Language (XML) 태그를 어떻게 처리하는지에 따라 토큰 수가 달라질 수 있기 때문이다. XML 태그는 발화 단위, 화자 및 담화 구조와 같은 정보를 표시하는 표지로서 실제 발화 내용과 구분되므로, 일반적으로 어휘 분석을 위한 토큰 수에서는 제외한다. 따라서, 공식 규모는 원 코퍼스의 사양을 설명하는 수치이며, 실제 분석에서는 연구자가 사용한 판본과 분석 프로그램에서 산출된 파일 수와 토큰 수를 별도로 보고해야 한다.

2.2. 전사 체계와 판본별 차이

Nesi와 Thompson (2006)에 의하면, BASE의 원 전사 파일은 Text Encoding Initiative (TEI) 지침을 적용한 XML 형식으로 제작되었다. TEI 지침은 전사 자료의 내용뿐 아니라 문서 구조, 담화 상황, 참여자 정보 및 비언어적 현상을 표준화된 태그와 속성으로 표시하기 위한 텍스트 부호화 체계이다. BASE 파일의 헤더에는 담화 사건과 참여자에 관한 메타데이터가 수록되어 있고, 본문에서는 발화 단위와 화자 정보가 구분된다. 일정 길이 이상의 휴지, 웃음, 기침, 중첩 발화와 같은 음성 현상뿐 아니라 몸짓, 판서 및 슬라이드 전환과 같은 비언어적 행동도 일부 표지되어 있다. 이러한 정보는 단어 빈도뿐 아니라 발화 전환, 상호작용 구조 및 시청각 자료의 사용을 연구할 수 있게 한다. 파일명에는 학문 분야, 담화 유형 및 개별 사건을 식별하는 정보가 포함된다. ‘ah’, ‘ls’, ‘ps’, ‘ss’는 각각 인문예술, 생명/의학, 물리과학 및 사회과학을 뜻하며, ‘lct’와 ‘sem’은 강의와 세미나를 나타낸다. 파일 코드가 보존된 판본에서는 이러한 정보를 이용하여 분야별, 담화 유형별 하위 코퍼스를 구성할 수 있다. 다만 모든 담화에 동일한 수준의 전사 표지가 부착된 것은 아니다. 강의는 일반적으로 한 명의 교수자가 발화를 주도하지만, 세미나는 여러 화자가 발언 순서, 억양, 시선과 몸짓을 이용하여 상호작용을 조절한다. 이 때문에 세미나 자료에는 강의보다 세부적인 비언어 정보가 덜 일관되게 표지되어 있다. 강의와 세미나에서 특정 태그의 빈도가 다르더라도 이를 실제 행동 빈도의 차이로 곧바로 해석해서는 안 된다.

전사 규약은 검색 결과에도 직접적인 영향을 준다. 예를 들어 BASE에서는 *gonna*, *wanna*, *kinda*와 같은 축약 발음을 각각 *going to*, *want to*, *kind of*로 풀어 적었다. 따라서 AntConc에서 *gonna*가 검색되지 않는다고 해서 실제 강의에서 해당 발음이 전혀 사용되지 않았다고 결론 내릴 수 없다. 전사문은 음성을 그대로 복제한 자료가 아니라 연구 목적과 표기 원칙에 따라 음성을 문자로 재구성한 자료이기 때문이다. BASE의 원자료, 연구용 처리본

및 온라인 플랫폼 가공판 사이에도 차이가 있다. 공식 원자료는 강의 160개와 세미나 39개, 1,644,942토큰으로 구성된다. Dang과 Webb(2014)이 사용한 처리본은 1,691,997토큰으로 보고되었으며, Sketch Engine 판은 중복되는 세미나 파일인 *Issem003*을 제외하여 강의 160개와 세미나 38개로 구성된다(Sketch Engine, 2025). 판본 간 수치 차이는 오류라기보다 중복 자료, 비언어 표지, 품사 태그, 문장부호와 토큰화 규칙을 다르게 처리한 결과이다. 따라서 연구의 재현성을 확보하려면 코퍼스 이름뿐 아니라 배포처, 판본, 접근일, 실제 파일 수, 토큰 수, 주석 여부 및 전처리 조건을 밝혀야 한다.

2.3. BASE 를 활용한 대표 연구

Nesi와 Basturkmen(2006)은 BASE 강의 120개와 MICASE 강의 40개를 이용하여 어휘다발이 대학 강의를 조직하는 방식을 분석하였다. 연구 결과, 빈도가 높은 단언어 표현의 상당수는 강의의 관계와 진행 방향을 알려 주는 담화 신호 기능을 수행하였다. *If you look at*이나 *in terms of the*와 같은 표현은 개별 단어의 의미만 보면 기능이 분명하지 않지만, 실제 강의에서는 주제 전환, 예시 제시 및 논리 관계를 안내하는 역할을 하였다. 이는 반복 표현을 단순한 군더더기가 아니라 청자가 강의 구조를 따라가도록 돕는 담화 자원으로 이해해야 함을 보여 준다. Grant(2011)는 BASE의 강의와 세미나를 표집하여 *just*의 빈도와 화용적 기능을 분석하였다. *Just*는 네 학문 분야 모두에서 최소화 또는 완화 기능으로 가장 자주 사용되었으며, 교수자의 설명, 지시와 요청을 부드럽게 만드는 역할을 하였다. 이 결과는 학습자가 흔히 ‘단지’ 또는 ‘방금’으로 이해하는 기능어가 실제 학술 구어에서는 설명의 강도와 대인 관계를 조절하는 복합적인 담화 자원으로 사용된다는 사실을 보여 준다.

Deroey와 Taverniers(2012)는 BASE의 강의 160개 전체를 분석하여 교수자가 중요한 내용을 어떻게 표시하는지 살펴보았다. 분석 결과, 785개의 관련성 표지가 확인되었으며, 이는 1만 단어당 약 7건에 해당하였다. *Remember...* 와 같은 동사 구문이 가장 두드러졌고, *the point is...* 와 같은 명사 구문이 그 뒤를 이었다. 교수자는 반드시 *This is important*라고 직접 말하지 않더라도 다양한 동사, 명사, 형용사, 부사 구문을 통해 핵심 개념과 강의 구조의 중요성을 표시하였다. Dang과 Webb(2014)은 BASE의 강의 160개와 세미나 39개 전체를 처리하여 학술 구어의 어휘 구성을 분석하였다. 고유명사와 주변어를 포함할 경우 상위 4,000어족은 전체 자료의 96.05%를, 상위 8,000어족은 98%를 포괄하였다. 분야별 95% 포괄에 필요한 어휘는 약 3,000~5,000어족, 98% 포괄에는 약 5,000~13,000어족으로 차이가 나타났다. 이는 대학 강의가 전문용어만으로 이루어지는 것이 아니라 고빈도 일반어가 학술 구어 이해의 중요한 토대를 형성한다는 점을 보여 준다. 다만 어휘 포괄률은 실제 듣기 이해율과 동일하지 않으며, 배경지식, 발음, 발화 속도 및 담화 구조도 강의 이해에 영향을 준다.

Wei, Leung과 Sun(2023)은 BASE의 세미나 9개에서 확인된 부가의문문 122개를 분석하였다. 전체 사례 가운데 81.15%를 교수자가 사용했으며, 73.77%는 실제 정보를 요구하기보다 내용을 강조하는 수사적 기능을 수행하였다. 순수한 정보 요청은 한 건에 불과했고, 상당수의 부가의문문 뒤에는 명시적인 구두 응답이 나타나지 않았다. 이는 부가의문문이 단순히 청자의 동의를 구하는 문법 형식이 아니라 친밀감을 형성하면서 교수자가 발언권과 수업의 진행 방향을 유지하는 상호작용 자원이 될 수 있음을 보여 준다. 이들 연구는 BASE가 학술 어휘의 빈도를 조사하는 자료에 그치지 않음을 보여 준다. 어휘다발, 기능어, 관련성 표지, 어휘 부담 및 부가의문문에 관한 연구를 통해 학술 구어의 어휘적 특성, 강의 조직, 대인 관계와 교수자 및 학생 상호작용을 통합적으로 분석할 수 있다. 동시에 연구마다 사용한 하위 자료, 표집 규모 및 처리 방식이 다르므로 결과를 비교할 때에는 BASE 전체를 분석했는지, 일부 강의나 세미나만을 분석했는지 구분해야 한다.

2.4. BASE 의 활용 가치와 해석상의 한계

BASE의 주요 장점은 네 개의 넓은 학문 분야가 거의 같은 수의 담화 사건으로 구성되어 있다는 점이다. 강의와 세미나를 함께 포함하기 때문에 교수자의 설명이 중심인 담화와 다수 참여자의 상호작용이 중심인 담화를 비교할 수 있다. 화자, 휴지, 중첩 발화와 비언어 행동에 관한 일부 정보도 제공되므로 단어 빈도를 넘어 담화와 상호작용을 분석할 수 있다. MICASE와의 비교를 통해 영국과 미국 대학의 학술 구어를 탐색할 수 있다는 점도 중요한 가치이다. 그러나 자료가 1998년부터 2005년까지 영국의 두 대학에서 수집되었다는 점은 일반화의 범위를 제한한다. BASE의 결과를 오늘날 모든 영국 대학, 온라인 및 하이브리드 수업, 다른 영어권 대학 또는 아시아 지역의 영어매개수업에 그대로 적용해서는 안 된다. 세미나 자료는 39개에 불과하고 물리과학 분야의 세미나는 9개이므로 저빈도 상호작용 현상을 분야별로 비교하기에는 표본이 충분하지 않을 수 있다. 학문 분야 역시 넓은 분류에 해당한다. 특정 표현이 물리과학 자료에서 상대적으로 자주 나타난다고 해서 모든 물리학자나 공학자가 해당 표현을

선호한다고 단정할 수 없다. 정규화 빈도와 함께 해당 표현이 출현한 파일 수, 강의자별 분포 및 실제 KWIC 문맥을 확인해야 한다. 한 명의 강의자가 특정 표현을 반복적으로 사용하여 전체 분야의 빈도를 높였을 가능성도 검토할 필요가 있다.

2.5. AntConc 를 활용한 교육적 적용

BASE는 AntConc를 활용한 자료 중심 학습에 유용하다. 첫 번째 활동에서는 강의와 세미나에서 *just*의 기능과 분포를 비교할 수 있다. 학습자는 KWIC 용례를 최소화, 완화, 시간, 초점 및 기타 기능으로 분류하고, 강의와 세미나의 빈도를 1만 토큰당 빈도로 정규화할 수 있다. Collocate와 Plot을 함께 사용하면 반복되는 결합 표현과 특정 파일에 대한 집중 여부도 확인할 수 있다. 두 번째 활동에서는 BASE와 MICASE의 강의 자료를 비교하여 영국과 미국 대학 강의에서 상대적으로 두드러지는 어휘와 담화 표현을 탐색할 수 있다. 다만 두 코퍼스의 전체 자료를 그대로 비교하기보다 담화 유형, 학문 분야 및 화자 역할이 유사한 강의 자료를 대응시켜야 한다. 핵심어 분석에서는 두 코퍼스의 빈도 차이가 우연으로 발생했다고 보기 어려운지를 나타내는 로그우도(log-likelihood)와, 그 차이의 방향 및 실질적인 크기를 보여주는 로그비(log ratio)를 함께 검토해야 한다. 로그우도 값이 크더라도 실제 빈도 차이는 작을 수 있으며, 로그비가 크더라도 출현 빈도가 지나치게 낮으면 결과가 불안정할 수 있다. 따라서 최소 빈도와 정규화 빈도를 함께 확인하고, 고유명사, 전사 태그 및 강의 주제어를 실제 담화적 기능을 수행하는 표현과 구분하여 해석할 필요가 있다.

세 번째 활동에서는 교수자가 중요한 내용을 표시하는 관련성 표지를 분석할 수 있다. 학습자는 *remember*, *important*, *point*, *note* 등을 검색한 뒤, 실제 관련성 표지에 해당하는 용례를 동사, 명사, 형용사, 부사 구문으로 분류할 수 있다. 두 명의 학습자가 동일한 용례를 독립적으로 코딩하고 일치도를 계산하게 하면 검색 가능한 개념의 조작화, 문맥에 근거한 기능 분류 및 코더 간 판단 차이를 함께 학습할 수 있다. 모든 활동에서는 사용한 BASE 판본, AntConc 버전, 파일 수, 토큰 수, 검색식, 대소문자와 토큰 정의, 태그 처리 방식 및 정규화 단위를 기록해야 한다. 이러한 절차는 학습자가 검색 결과를 산출하는 데 그치지 않고, 자료의 대표성, 분석의 재현성 및 해석의 한계를 체계적으로 이해하게 한다.

3. 결론

본고는 BASE의 구축 배경과 자료 구성, 전사 체계와 판본별 차이를 검토하고, 대표 연구를 통해 학술적 활용 가치를 살펴보았다. BASE는 1998년부터 2005년까지 영국의 두 대학에서 녹음된 강의와 세미나를 체계적으로 수집한 학술 구어 코퍼스이다. 네 개의 넓은 학문 분야를 비교적 고르게 포함하고, 강의와 세미나를 함께 제공한다는 점에서 학문 분야와 담화 유형에 따른 학술 구어의 변이를 연구하는 데 유용하다. 대표 연구들은 반복되는 어휘다발이 강의의 흐름을 안내하고, *just*와 같은 기능어가 설명과 지시의 강도를 조절하며, 교수자가 다양한 문법 구문으로 중요 내용을 표시한다는 사실을 보여 주었다. 또한 대학 강의 이해의 토대를 이루는 것은 전문용어뿐 아니라 고빈도 일반어이며, 부가의문문도 정보 요청보다는 강조와 상호작용 조절 기능으로 사용될 수 있음이 확인되었다. 이러한 결과는 학술 구어 연구가 단어 빈도뿐 아니라 담화 조직, 대인 관계, 교수자와 학생 상호작용을 함께 고려해야 함을 시사한다.

BASE를 사용할 때에는 원자료와 가공판의 차이에 주의해야 한다. 공식 원자료, 연구용 처리본 및 Sketch Engine 버전(Sketch Engine, 2025)은 파일 수, 토큰 수와 품사 주석 방식이 서로 다르다. 전사 규약과 태그의 밀도도 분석 결과에 영향을 줄 수 있다. 따라서 연구자는 실제 사용한 판본, 파일 범위, 접근일, 토큰화 및 태그 처리 조건을 명시해야 한다. 이러한 정보가 누락되면 동일한 코퍼스 이름을 사용하더라도 연구 결과를 정확하게 재현하기 어렵다. BASE의 결과는 1998년부터 2005년까지 영국의 두 대학에서 생산된 강의와 세미나의 특성으로 제한하여 해석하는 것이 타당하다. 이를 현대 영국 대학 전체, 모든 영어권 대학 또는 온라인 수업으로 확대하기 위해서는 시기, 지역, 대학, 학문 분야 및 수업 유형이 다른 후속 코퍼스와 비교가 필요하다. 특히 오늘날의 영어매개수업과 온라인 및 하이브리드 강의를 수집한 자료를 구축하여 BASE와 비교한다면 대학 학술 구어의 변화 양상을 더욱 구체적으로 규명할 수 있을 것이다.

교육적 측면에서 BASE는 학습자가 실제 대학 강의와 세미나의 언어를 직접 관찰하도록 한다. AntConc를 활용한 기능어, 핵심어 및 관련성 표지 분석은 학습자가 사전적 의미나 문법 규칙만으로 설명하기 어려운 학술 담화의 기능을 자료에 근거하여 발견하게 한다. 따라서 BASE는 영국 대학 학술 구어의 역사적 기록일 뿐 아니라

코퍼스 분석의 재현성, 담화 기능의 문맥 의존성 및 학술 영어교육의 실제성을 함께 가르칠 수 있는 연구 및 교육 자원으로 평가할 수 있다.

참고문헌

- Anthony, L. (2026). *AntConc* (Version 4.4.2) [Computer software]. Tokyo, Japan: Waseda University. Retrieved July 20, 2026, from <https://www.laurenceanthony.net/software/AntConc>
- Corpus Resource Database. (2014). British Academic Spoken English Corpus (BASE). VARIENG, University of Helsinki. Retrieved June 2, 2026, from <https://varieng.helsinki.fi/CoRD/corpora/BASE/>
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66-76. <https://doi.org/10.1016/j.esp.2013.08.001>
- Deroey, K. L. B., & Taverniers, M. (2012). “Just remember this”: Lexicogrammatical relevance markers in lectures. *English for Specific Purposes*, 31(4), 221-233.
- Grant, L. E. (2011). The frequency and functions of just in British academic spoken English. *Journal of English for Academic Purposes*, 10(3), 183-197.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), 283-304.
- Nesi, H., & Thompson, P. (2006). The British Academic Spoken English Corpus manual. Universities of Warwick and Reading. Retrieved June 2, 2026, from https://warwick.ac.uk/fac/soc/al/research/collections/base/history/base_manual.pdf
- Sketch Engine. (2025). British Academic Spoken English corpus (BASE). Retrieved August 7, 2026, from <https://www.sketchengine.eu/british-academic-spoken-english-corpus/>
- Thompson, P., & Nesi, H. (2001). The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research*, 5(3), 263-264.
- University of Michigan English Language Institute. (2026). The Michigan corpora: MICASE + MICUSP. Retrieved July 7, 2026, from <https://lsa.umich.edu/eli/language-resources/micase-micusp.html>
- Wei, L., Leung, A. H.-C., & Sun, Y. (2023). “We don’t normally require that in other contexts, do we”: Interpersonal meanings of tag questions in British university seminars based on the BASE corpus. *Frontiers in Psychology*, 13, Article 1070937. <https://doi.org/10.3389/fpsyg.2022.1070937>

THE AUTHORS

Young Ok Jong is a Visiting Research Fellow at the Institute of Foreign Language Education, Hankuk University of Foreign Studies, South Korea. She holds an MSc in English Language Teaching from the University of Warwick and an MSc in Child Development and Education from the University of Oxford. She earned her EdD in Applied Linguistics and English Language Teaching from the University of Warwick. She is the author of *Effective Approaches to Teaching English Writing: A Hybrid English Writing Model* (2013). Her research interests include corpus linguistics, genre-based writing pedagogy, and process approaches to writing instruction.

Chae Kwan Jung is an Associate Professor in the Department of English Language and Literature at Incheon National University, South Korea. He holds a BEng (Hons) in Manufacturing Engineering and Japanese from the University of Birmingham, UK, and an MSc in Engineering Business Management and an EdD in Applied Linguistics and English Language Teaching from the University of Warwick, UK. He is the founder of the Institute for Corpus Research (ICR) and the founding editor of the Asia Pacific Journal of Corpus Research (APJCR). His research interests include corpus linguistics, English language education, English for Specific Purposes (ESP), language assessment, and curriculum design and development. His current interests include data-driven learning, AI-assisted language pedagogy and translation, and the application of ESP-informed genre- and corpus-based approaches to language teaching and learning.

THE AUTHORS' ADDRESSES**First Author****Young Ok Jong**

Visiting Research Fellow

Institute of Foreign Language Education

Hankuk University of Foreign Studies

107 Imun-ro, Dongdaemun-gu, Seoul, 02450, SOUTH KOREA

E-mail: yojong@gmail.com

Corresponding Author**Chae Kwan Jung**

Professor

Department of English Language & Literature

Incheon National University

119 Academy-ro, Yeonsu-gu, Incheon, 22012, SOUTH KOREA

E-mail: ckjung@inu.ac.kr

Received: 13 June 2026

Received in Revised Form: 13 July 2026

Accepted: 15 August 2026