

영국 학술 문어 코퍼스(BAWE): 구축, 연구 활용 및 AntConc 활용 교육

김보민
(한국외국어대학교)
정채관
(국립인천대학교)

Kim, B. M., & Jung, C. K. (2026). The British Academic Written English Corpus (BAWE): Compilation, research applications, and AntConc-based pedagogy. *Asia Pacific Journal of Corpus Research*, 7(1), 13-19.

This article reexamines the British Academic Written English Corpus (BAWE) as a resource for investigating university assignments and supporting corpus-based writing education. Drawing on corpus documentation and five studies, it reviews BAWE's compilation, stratified sampling, analytical units, genre classification, metadata, versions, research applications, and pedagogical potential. BAWE contains 2,761 assignments and 6,506,995 words produced in the early 2000s by undergraduate and taught master's students at several British universities. The review shows that university writing extends well beyond essays to thirteen genre families and that disciplinary and genre differences may outweigh level-related variation. Research using BAWE indicates that advanced writing may rely more on complex noun phrases than subordinate clauses, that lexical bundles reveal disciplinary relationships, and that successful writers with different first-language backgrounds use cohesive resources in varied ways. Strengths include authentic assessed writing and rich contextual metadata. However, differences among assignments, independent texts, genre units, and corpus versions complicate replication. Selection of higher-rated work, uneven subgroup sizes, and limited temporal and institutional coverage constrain generalization. Three AntConc activities are proposed: comparing first-person expressions across genres, contrasting BAWE with MICUSP, and developing discipline-specific writing guides. BAWE remains valuable when findings are interpreted within genre, discipline, level, and assessment contexts.

Keywords: British Academic Written English Corpus (BAWE), University Student Writing, Academic Writing Genres, AntConc, Data-Driven Learning

1. 서론

대학의 학술 글쓰기 교육은 오랫동안 출판된 연구 논문이나 학위논문을 모범으로 삼아 이루어져 왔다. 그러나 학부생과 수업형 석사과정 학생이 작성하는 평가 과제는 전문 연구자가 학문 공동체에 새로운 지식을 보고하기 위해 작성하는 논문과 목적, 독자 및 평가 상황이 다르다. 학생의 과제는 대체로 정해진 독자인 전공 교원에게 자신의 지식, 분석 능력과 전문적 판단을 입증하기 위한 글이다. 따라서 대학생에게 필요한 학술 글쓰기를 이해하려면 전문가의 출판물뿐 아니라 실제 교과목에서 제출되고 평가된 학생의 글을 직접 분석할 필요가 있다.

British Academic Written English Corpus(이하, 'BAWE')는 이러한 필요에 따라 영국 대학생이 정규 교과목에서 작성하고 전공 교원에게 좋은 평가를 받은 과제를 수집하여 구축한 학술 문어 코퍼스이다(Coventry University, 2025). BAWE는 영국 경제사회연구위원회의 지원을 받아 University of Warwick, University of Reading과 Oxford Brookes University가 공동으로 구축하였으며, 표집이 부족한 분야를 보완하기 위해 프로젝트 마지막 단계에 Coventry University의 자료도 포함하였다. 과제 수집은 2004년부터 2007년까지 이루어졌고, 수록 자료는 주로 2000년대 초반에 작성되었다(Alsop & Nesi, 2009; Nesi, 2008). BAWE의 공식 규모는 과제 2,761편과 6,506,995 단어이며, 학부 1학년부터 수업형 석사과정까지 4개 학업 수준과 35개 학문 분야를 포함한다. 자료는 예술/인문학, 생명과학, 물리과학 및 사회과학이라는 네 개의 넓은 학문 분야군으로 구성된다. 에세이뿐 아니라 사례 연구, 비평, 실험 보고서, 설계 명세, 제안서, 연구 보고서와 성찰적 서사 등 다양한

평가 과제를 수록하고 있다는 점도 중요한 특징이다.

BAWE에서 말하는 ‘성공적인 글쓰기’, 즉 *proficient writing*은 표준화된 영어시험에서 높은 점수를 받은 사람이 작성한 글이나 언어적 오류가 없는 글을 의미하지 않는다. 이는 영국 대학의 정규 교과목에서 실제로 제출된 과제 가운데 60점 이상을 받아 해당 전공, 과제 장르 및 학업 단계의 평가 기준을 충분히 충족한 것으로 인정된 글을 가리킨다. 여기서 숙련성은 일반적인 영어 능력에 대한 절대적 판단이 아니라, 과제의 목적에 적절히 대응하고 전공 지식, 논증 방식, 자료 활용 및 장르 관습을 성공적으로 구현했는지에 관한 맥락 의존적 평가이다. 따라서 BAWE는 학생들의 보편적인 영어 숙달도를 대표하는 코퍼스라기보다, 실제 학문 공동체의 평가 관행에서 양호하거나 우수한 성취를 인정받은 학생 과제를 수집한 코퍼스로 이해해야 한다(Nesi, 2008).

BAWE를 영국 원어민 대학생의 모범 영어 코퍼스로 규정하는 것도 적절하지 않다. 영어를 제1언어로 사용하는 학생이 다수를 이루지만, 그 밖의 제1언어 배경을 지닌 학생의 과제도 포함되어 있다. 자료 선정 기준은 학생의 출신이나 모어가 아니라 과제가 전공의 평가 기준을 충족했는가였다. BAWE는 오류를 중심으로 구축한 전형적인 학습자 코퍼스와 구별되지만, 수록된 모든 문장을 보편적으로 따라야 할 규범으로 제시하는 코퍼스도 아니다. 성공적인 글쓰기의 특성 역시 장르, 전공, 학업 수준과 과제 조건 안에서 해석해야 한다. 본고의 목적은 BAWE의 구축 배경과 표집 설계, 자료 구성, 장르 분류 및 메타데이터를 검토하고, 대표적인 연구 사례를 통해 이 코퍼스의 학술적·교육적 활용 가치를 고찰하는 데 있다. 아울러 과제·텍스트·장르 분석 단위의 차이, 판본별 자료 구성, 분석 결과의 일반화 범위와 이용 조건을 살펴보고, AntConc(Anthony, 2026)를 활용한 연구·교육 활동을 제안하고자 한다.

2. British Academic Written English (BAWE)

2.1. BAWE의 구성과 표집 설계

Nesi, Sharpling과 Ganobcsik-Williams(2004)에 따르면, BAWE의 구축에 앞서 2001년에 조성된 예비 코퍼스에는 70명의 학생이 제출한 과제 499편이 포함되었다. 그러나 자발적으로 제공받은 자료는 인문/사회계열과 학부 1학년 과제에 편중되었고, 자연과학 및 상위 학업 수준의 자료는 상대적으로 부족하였다. 이 경험은 자발적인 자료 수집만으로 전공과 학업 수준의 균형이 자연스럽게 확보되지 않는다는 점을 보여 주었다. 본 프로젝트에서는 이러한 문제를 보완하기 위해 네 개 학문 분야군과 네 개 학업 수준을 결합한 4×4 층화 표집을 적용하였다. 각 표집 칸에 비슷한 수의 과제를 포함하고, 특정 필자나 교과목이 전체 자료를 지나치게 지배하지 않도록 자료를 수집하였다. 과제가 잘 확보되지 않는 분야에서는 사례비를 높이고, 추가 대학에서 자료를 수집하는 방식으로 부족한 영역을 보완하였다(Alsop & Nesi, 2009).

이러한 표집 설계는 당시 영국 대학에서 실제로 생산된 글의 양을 비례적으로 재현하기보다 학문 분야와 학업 수준을 비교할 수 있는 분석 가능성을 우선한 것이다. 따라서 BAWE의 분야별 과제 수가 비슷하다는 사실을 영국 대학에서 네 분야의 글이 실제로 비슷한 양으로 생산되었다는 의미로 해석해서는 안 된다. 대표성은 코퍼스의 규모뿐 아니라 표집 목적, 모집 경로, 참여 동기와 분야별 보완 절차를 함께 고려하여 판단해야 한다(Jung & Wharton, 2012). 공식 자료에서 예술/인문학은 705개 과제와 1,714,118 단어, 생명과학은 683개 과제와 1,412,391 단어, 물리과학은 596개 과제와 1,381,356 단어, 사회과학은 777개 과제와 1,999,130 단어로 구성된다. 학업 수준별로는 Level 1이 807개, Level 2가 767개, Level 3이 591개, Level 4가 587개이며, 수준 정보가 없는 과제도 9개 포함되어 있다. Level 4는 박사과정이 아니라 수업형 석사과정에 해당한다.

BAWE는 동일한 학생을 여러 해 추적한 종단 코퍼스가 아니다. 학업 수준별 자료는 서로 다른 학생 집단이 작성한 과제로 구성된다. 따라서 Level 1과 Level 4의 차이를 한 학생이 시간이 지나면서 변화한 결과로 해석해서는 안 된다. 학업 수준에 따른 차이는 전공, 장르, 학생 구성 및 과제 조건의 영향과 함께 검토해야 한다.

2.2. 과제/텍스트/장르 분석 단위와 메타데이터

Heuboeck, Holmes와 Nesi(2010)에 의하면, BAWE의 기본 수집 단위는 학생이 한 번의 평가를 위해 제출한 과제이다. 그러나 하나의 과제 안에 실험 보고서와 성찰문처럼 서로 다른 목적을 지닌 독립적인 글이 포함될 수 있다. 이러한 복합 과제는 평가 단위로는 하나이지만, 장르 분석에서는 둘 이상의 텍스트로 분리될 수 있다. 그러므로 과제 수, 파일 수, 독립 텍스트 수 및 장르 분석 단위는 서로 일치하지 않을 수 있다. 공식적으로 가장 안정적으로 확인되는 수치는 과제 2,761편과 6,506,995 단어이다. 반면 독립 텍스트 수는 자료 정리 시점과 분할 기준에 따라 2,896편

또는 2,897편으로 보고되었다. 장르군 분포표의 합계는 2,859개이며, Gardner와 Nesi(2013)는 2,858개 텍스트를 분석한 반면, Gardner, Nesi와 Biber(2019)는 메타데이터 조건을 충족하는 2,760개 과제를 사용하였다.

이러한 차이는 특정 수치가 반드시 잘못되었다는 뜻이 아니라 연구 단위와 포함 기준이 서로 달랐음을 보여 준다. 연구자는 과제를 하나의 제출 및 평가 단위로 분석했는지, 복합 과제를 독립 텍스트로 분리했는지, 장르 정보가 있는 자료만 포함했는지, 결측 메타데이터를 제외했는지를 구체적으로 밝혀야 한다. 코퍼스 규모는 숫자만으로 제시하기보다 분석 단위의 정의와 함께 보고해야 재현성을 확보할 수 있다. 한편, BAWE 연구진은 과제 지시문에 사용된 명칭만을 따르지 않고 실제 제출물의 사회적 목적과 전개 구조를 분석하여 자료를 13개 장르군으로 분류하였다. 장르군에는 사례 연구, 비평, 설계 명세, 공감적 글쓰기, 에세이, 연습 과제, 설명문, 문헌 조사, 방법론 수행 보고, 서사적 회고, 문제형 과제, 제안서 및 연구 보고서가 포함된다(Gardner & Nesi, 2013; Nesi & Gardner, 2012).

BAWE 파일은 Extensible Markup Language(이하, 'XML') 형식으로 부호화되어 있다. XML은 과제 본문과 그 과제에 관한 배경정보를 태그로 구분하여 컴퓨터가 처리할 수 있게 하는 문서 형식이다. 각 파일의 XML 헤더에는 학업 수준, 작성 시기, 교과목명, 학문 분야군, 세부 전공, 장르군, 성적 및 과제 작성에 참여한 학생 수와 같은 메타데이터가 기록되어 있으며, XML 텍스트에는 학생이 작성한 실제 과제 본문이 수록되어 있다. 학생이 제공에 동의한 성별, 출생 연도, 제1언어 및 영국 중등교육 경험 등의 배경정보도 개인을 직접 식별할 수 없도록 처리되어 포함된다. 이러한 구조를 활용하면 특정 학업 단계, 전공, 장르 또는 학생 배경에 해당하는 자료를 선별하여 하위 코퍼스를 구성할 수 있다. 다만 모든 메타데이터 항목이 모든 파일에 완전하게 기록된 것은 아니므로, 하위 집단을 비교할 때에는 결측값과 범주별 파일 수, 작성자 수, 토큰 수를 먼저 확인해야 한다.

2.3. 판본과 자료 형식에 따른 분석 조건

BAWE는 Oxford Text Archive(이하, 'OTA')의 XML판, 일반 텍스트판, BAWE Plus의 PDF 자료 및 온라인 검색/파생 분석판 등 여러 형식으로 제공된다. BAWE Plus는 정식 BAWE 코퍼스에 텍스트를 추가한 하나의 확장 코퍼스라기보다, 21세기 영국 대학의 학술 문어를 다각적으로 연구하기 위해 워릭대학교가 보관하는 관련 자료의 집합이다. 여기에는 BAWE 코퍼스 본체, 과제의 원래 문서 구조를 확인할 수 있는 PDF 파일, 2001년에 구축된 약 100만 단어 규모의 BAWE 예비 코퍼스, 참여 학과 교원과의 면담 기록 및 Warwick English Language Test의 우수 서면 답안 자료가 포함된다. 이러한 자료를 활용하면 학생 과제의 언어적 특징을 문서 구성, 전공별 평가 관행 및 영어 능력 평가 자료와 연결하여 분석할 수 있다(University of Warwick, 2025).

XML판에는 구조 표지와 과제별 메타데이터가 보존되어 있고, 일반 텍스트판은 AntConc와 같은 프로그램에서 본문을 검색하기에 편리하다. BAWE Plus의 PDF 자료는 익명화된 원문서의 배치, 표, 목록 및 시각적 구성을 확인하는데 유용하다. 온라인 파생판에는 플랫폼에 따라 토큰화, 품사 주석과 구문 분석 정보가 추가되어 있을 수 있다. 이들 판본은 동일한 BAWE에서 출발하지만 보존하는 정보와 토큰 수가 같지 않다. 일반 텍스트판에서는 XML 헤더의 메타데이터와 원문서의 시각적 배치가 누락될 수 있다. PDF 문서는 언어 빈도 분석에 사용하는 텍스트 파일과 동일한 분석 단위로 간주하기 어렵다. 온라인 파생판의 토큰 수 역시 공식 규모인 6,506,995 단어와 다를 수 있다. AntConc에서 BAWE를 분석하려면 사용한 판본, 파일 수, 토큰 수와 타입 수를 먼저 기록해야 한다. XML 태그, 참고문헌, 표/그림 표지, 수식 및 URL 대체 문자열을 검색 대상에 포함했는지도 밝혀야 한다. 일반 텍스트 파일만 불러오면 장르, 전공과 학업 수준이 자동으로 표시되지 않을 수 있으므로 BAWE 메타데이터 파일과 개별 텍스트 파일을 연결하여 하위 코퍼스를 구성해야 한다.

2.4. BAWE 를 활용한 대표 연구

Gardner와 Nesi(2013)는 BAWE의 실제 학생 과제 2,858개를 분석하여 대학 평가 글쓰기의 13개 장르군을 제안하였다. 연구진은 과제 명칭이나 교원의 직관에만 의존하지 않고, 실제 텍스트의 구조, 학생과 교원의 면담 및 학과 자료에서 확인한 교육적 목적을 종합하였다. 분석 결과, 대학의 평가 글쓰기는 에세이에 한정되지 않았으며, 사례 분석, 연구 능력, 전문 실무 준비, 독립적 추론 및 성찰을 요구하는 다양한 장르로 구성되어 있었다. 이는 학술 글쓰기 교육이 일반적인 에세이 구조뿐 아니라 전공별 과제의 구체적인 사회적 목적과 전개 방식을 다루어야 함을 시사한다. Gardner, Nesi와 Biber(2019)는 BAWE 과제 2,760편을 대상으로 다차원 분석을 실시하여 학문 분야, 장르 및 학업 수준이 약 150개의 어휘와 문법 특징에 미치는 영향을 비교하였다. 일부 언어 차원에서는 장르와 학문 분야가 변이의 상당 부분을 설명한 반면, 학업 수준의 설명력은 상대적으로 낮았다. 다른 차원에서는 학문 분야와

학업 수준의 설명력이 비슷하게 나타났다. 이 결과는 고학년의 글이 모든 측면에서 일관되게 더 학술적으로 변화하는 것이 아니며, 학생이 어느 수준에 있는가뿐 아니라 무엇을 어떤 전공과 장르의 관습 안에서 쓰는가가 중요함을 보여 준다.

Staples, Egbert, Biber와 Gray(2016)는 BAWE에서 영어를 제1언어로 보고한 학생의 과제를 활용하여 학업 수준, 학문 분야 및 장르에 따른 문법적 복잡성을 분석하였다. 전반적으로 학업 수준이 높아질수록 명사구와 전치사구를 활용한 구 단위 복잡성이 증가한 반면, 일부 절 단위 복잡성은 감소하였다. 다만 변화의 방향과 크기는 전공과 장르에 따라 달랐다. 이는 종속절이 많고 문장이 길수록 발전된 글이라는 단순한 관점에 의문을 제기하며, 고급 학술 글쓰기에서는 정보를 명사구 안에 조밀하게 압축하는 능력이 중요할 수 있음을 보여 준다. Durrant(2017)는 BAWE에 나타난 네 단어 어휘 묶음을 분석하여 전공 간 언어적 거리를 시각화하였다. 과학/공학 분야와 인문/사회 분야 사이에는 비교적 큰 차이가 나타났고, 생명과학과 상경 분야는 두 영역의 특성을 공유하는 중간적 위치를 보였다. 같은 전공 안에서도 학생 간 차이가 적지 않았다. 이 연구는 학문 분야를 단순히 인문계와 이공계로 구분하기보다, 서로 인접하거나 혼합적인 성격을 지닌 연속체로 탐색할 수 있음을 보여 준다.

Nesi와 Moreton(2012)은 영어를 제1언어로 보고한 필자의 과제와 그 밖의 제1언어 배경을 지닌 필자의 과제에서 껍질 명사(shell noun)의 사용을 비교하였다. 껍질 명사는 *fact*, *problem*, *method* 및 *result*처럼 문맥에서 분리하면 의미가 비교적 일반적이지만, 앞이나 뒤에 제시된 복잡한 내용을 하나의 개념으로 요약하고 분류하는 추상명사를 말한다. 예를 들어 *The sample was too small. This problem affected the results*에서 *this problem*은 앞 문장의 내용을 ‘문제’로 요약하면서 두 문장을 연결한다. 이러한 기능 때문에 껍질 명사는 문장 간 의미 관계를 형성하는 텍스트 결속 장치인 동시에, 해당 내용을 사실, 문제, 원인, 결과 등으로 평가하는 태도 표현으로도 기능한다. 분석 결과, 영어 이외의 제1언어 집단이 이러한 명사를 일관되게 적게 사용한다는 증거는 나타나지 않았다. 일부 명사는 오히려 해당 집단에서 더 높은 빈도를 보였지만, 이는 과제가 특정 학문 분야에 더 많이 분포한 결과일 수 있었다. 이는 관찰된 차이를 제1언어나 영어 능력의 영향만으로 설명하지 않고 장르, 전공 및 학업 수준을 함께 통제해야 한다는 점을 보여 준다.

이상의 연구를 종합하면 BAWE의 핵심 가치는 학생 언어의 전체 빈도를 제시하는 데 있지 않다. 풍부한 메타데이터를 이용하여 언어적 특성을 장르, 전공, 학업 수준, 성적 및 필자의 언어 배경과 연결할 수 있다는 점에 있다. 동시에 어느 요인이 가장 중요한지는 분석하는 언어 특징과 연구 질문에 따라 달라질 수 있으므로 하나의 변인만으로 학생 글쓰기의 차이를 단정하는 것은 위험할 수 있다.

2.5. BAWE의 활용 가치와 해석상의 한계

BAWE는 실제 교과목에 제출되어 전공 교원에게 평가받은 과제를 수록한다는 점에서 높은 생태적 타당도를 지닌다. 짧은 시험 에세이나 연구자가 요구한 통제 과제가 아니라 학생이 실제 평가 상황에서 작성한 글을 제공한다. 학문 분야, 학업 수준, 장르, 성적과 학생의 언어 배경에 관한 메타데이터가 풍부하며, 성공적으로 평가된 전공별 글쓰기의 관습을 탐색하기에 적합하다. 그러나 BAWE에는 60점 이상의 과제만 포함되어 있어 낮은 성적의 과제, 초고, 수정 과정과 교원의 피드백을 연구하기 어렵다. 자발적으로 제출된 자료를 층화 표집으로 보완하였으므로 영국 대학 전체에서 실제로 생산되는 글의 양을 비례적으로 대표하지도 않는다. 일부 장르군은 35~40개에 불과하여 학문 분야와 학업 수준까지 세분하면 표본이 매우 작아질 수 있다. BAWE는 2000년대 초 영국의 일부 대학에서 수집된 자료이므로 최근의 온라인 평가, 생성형 AI 활용 및 변화한 대학 글쓰기 관행을 반영하지 않는다. 박사과정의 글과 학위논문보다 학부 및 수업형 석사과정의 교과목 과제에 초점을 둔다. 따라서 분석 결과를 현대 영국 대학 전체, 다른 국가의 대학 글쓰기, 학술 논문이나 학위논문으로 확대해서는 안 된다. BAWE에서 특정 표현이 자주 나타난다는 사실을 좋은 성적의 직접적인 원인으로 해석하는 것도 무리가 있다. 높은 빈도는 장르가 요구하는 절차, 특정 전공의 내용어, 과제 지시문, 인용 관행 또는 소수 필자의 반복에서 비롯될 수 있다. 원빈도뿐 아니라 정규화 빈도, 파일 범위, 필자별 분포와 실제 문맥을 함께 검토하고, 코퍼스에서 관찰된 경향과 교육적으로 권장할 규범을 구분해야 한다.

2.6. AntConc를 활용한 교육적 적용

첫 번째 활동에서는 에세이, 방법론 수행 보고 및 서사적 회고에 나타난 *I*, *we*, *my*, *our*의 빈도와 기능을 비교할 수 있다. 학습자는 각 장르의 하위 코퍼스를 구성하고 1인칭 표현의 빈도를 정규화한 뒤, 용례를 연구 절차 보고, 주장과 태도, 개인 경험, 독자 포함 및 인용 등으로 분류할 수 있다. 이 활동은 1인칭 표현의 빈도만으로 글의

주관성을 판단해서는 안 되며, 과제의 목적과 지시 대상을 함께 확인해야 함을 보여 준다. 두 번째 활동에서는 BAWE와 Michigan Corpus of Upper-level Student Papers(MICUSP)를 비교하여 영국과 미국의 성공적인 대학 과제에 나타난 태도·논증·필자 지칭 표현을 탐색할 수 있다. 두 코퍼스에서 가능한 한 유사한 전공, 장르와 학업 수준을 선택하고, 핵심어 목록과 KWIC 문맥을 분석해야 한다. 발견된 차이는 우선 선택된 두 하위 코퍼스의 차이로 기술하고, 국가 차이는 가능한 설명 가운데 하나로만 제시하는 것이 바람직하다.

세 번째 활동에서는 학습자가 자신의 전공과 장르에 적합한 자료 기반 학술 글쓰기 안내서를 만들 수 있다. 특정 전공과 장르의 자료를 대상 코퍼스로 설정하고, BAWE 전체 또는 다른 전공의 동일 장르를 참조 코퍼스로 사용하여 핵심어와 3~5단어 어휘 묶음을 추출한다. 이후 KWIC와 Plot을 이용하여 의미, 문법적 결합, 담화 기능 및 파일별 분포를 분석하고, 여러 과제에 고르게 나타난 표현을 중심으로 교육 자료를 구성할 수 있다. 모든 활동에서는 사용한 BAWE 판본, 하위 코퍼스 구성 기준, 파일 수, 토큰 수, 검색식, 참고문헌과 인용문의 포함 여부, 정규화 단위 및 파일 범위를 기록해야 한다. 원문 파일을 재배포하지 않고 이용 조건이 허용하는 범위에서 짧은 용례와 집계 결과만 제시하는 연구 윤리도 함께 지도할 필요가 있다.

3. 결론

본고는 BAWE의 구축 배경과 표집 설계, 자료 단위, 장르 분류, 메타데이터와 판본별 특성을 검토하고, 대표 연구를 통해 학술적, 교육적 활용 가능성을 살펴보았다. BAWE는 영국 대학의 학부 및 수업형 석사과정 학생이 실제 교과목에서 제출하여 성공적으로 평가받은 과제를 수집했다는 점에서 중요한 가치를 지닌다. 특히 에세이에 한정되지 않은 다양한 과제 장르와 전공, 수준, 성적, 언어 배경에 관한 메타데이터는 대학생 학술 글쓰기의 복합적인 변이를 분석할 수 있게 한다. 대표 연구들은 학생 글쓰기의 차이가 학업 수준만으로 설명되지 않으며, 장르와 학문 분야가 중요한 영향을 미친다는 사실을 보여 주었다. 발전된 학술 글쓰기에서는 절의 수보다 명사구와 전치사구를 이용한 정보 압축이 중요할 수 있고, 반복되는 어휘 묶음은 전공 간 유사성과 차이를 드러낼 수 있다. 또한 영어 이외의 제1언어 배경을 지닌 학생의 성공적인 글을 일률적으로 결속성이 낮거나 미숙한 글로 규정하기 어렵다는 점도 확인되었다.

BAWE를 연구에 사용할 때에는 과제, 독립 텍스트 및 장르 분석 단위를 구분해야 한다. 연구 목적에 따라 분석 자료의 총계가 달라질 수 있으므로 사용한 판본, 포함/제외 기준, 메타데이터 결측값과 복합 과제의 분할 방식을 구체적으로 보고해야 한다. 코퍼스에 자주 나타난 표현을 성공적인 글쓰기의 원인이나 보편적인 규범으로 해석하지 않고, 장르·전공·수준과 실제 문맥 안에서 설명하는 것도 중요하다. BAWE는 2000년대 초 영국 일부 대학의 성공적인 평가 과제를 보여 주는 코퍼스이다. 따라서 오늘날의 디지털 및 AI 매개 글쓰기, 낮은 성적의 글, 초고와 수정 과정, 박사과정의 학위논문 및 다른 국가의 대학 글쓰기까지 직접 대표하지 않는다. 향후에는 BAWE와 최근 대학 과제 코퍼스를 비교하고, 생성형 AI 도입 전후의 과제 장르, 문법적 복잡성, 인용 관행 및 필자 정체성 표현의 변화를 분석할 필요가 있다.

교육적 측면에서 BAWE는 추상적인 학술 글쓰기 원칙을 실제 전공 과제의 언어와 연결할 수 있는 자료이다. AntConc를 이용한 장르별 1인칭 표현 분석, BAWE와 MICUSP의 비교 및 전공별 글쓰기 안내서 개발은 학습자가 빈도, 분포와 문맥을 종합하여 판단하도록 한다. 이러한 점에서 BAWE는 성공적인 대학 과제의 언어적 특징을 기술하는 코퍼스일 뿐 아니라 장르와 전공에 민감한 학술 글쓰기 교육과 재현 가능한 코퍼스 분석을 지원하는 연구 및 교육 자료로 평가할 수 있다.

참고문헌

- Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71-83. <https://doi.org/10.3366/E1749503209000227>
- Anthony, L. (2026). *AntConc* (Version 4.4.2) [Computer software]. Tokyo, Japan: Waseda University. Retrieved July 22, 2026, from <https://www.laurenceanthony.net/software/AntConc>
- Coventry University. (2025). British Academic Written English Corpus (BAWE). Retrieved June 12, 2026, from <https://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/>

- Durrant, P. (2017). Lexical bundles and disciplinary variation in university students' writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193.
- Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics*, 34(1), 25-52.
- Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, 40(4), 646-674.
- Heuboeck, A., Holmes, J., & Nesi, H. (2010). The BAWE corpus manual (Version III). British Academic Written English Corpus Project. Retrieved June 12, 2026, from <https://www.coventry.ac.uk/globalassets/media/global/08-new-research-section/current-projects/bawemanual-v3.pdf>
- Jung, C. K., & Wharton, S. (2012). Finding textual examples of genres: Issues for corpus users. *Korean Journal of English Language and Linguistics*, 12(1), 129-148.
- Nesi, H. (2008). BAWE: An introduction to a new resource. In A. Frankenberg-Garcia (Ed.), *Proceedings of the 8th Teaching and Language Corpora Conference*. ISLA Lisboa. Retrieved July 11, 2026, from <https://warwick.ac.uk/fac/soc/al/research/collections/bawe/papers/talc8.pdf>
- Nesi, H., & Gardner, S. (2012). *Genres across the Disciplines: Student Writing in Higher Education*. Cambridge: Cambridge University Press.
- Nesi, H., & Moreton, E. (2012). EFL/ESL writers and the use of shell nouns. In R. Tang (Ed.), *Academic Writing in a Second or Foreign Language: Issues and Challenges Facing ESL/EFL Academic Writers in Higher Education Contexts* (pp. 126-145). London: Continuum.
- Nesi, H., Sharpling, G., & Ganobcsik-Williams, L. (2004). Student papers across the curriculum: Designing and developing a corpus of British student writing. *Computers and Composition*, 21(4), 439-450.
- Nesi, H., Gardner, S., Thompson, P., & Wickens, P. (2004). British Academic Written English Corpus [Data set]. Oxford Text Archive. Retrieved May 7, 2026, from <https://ota.bodleian.ox.ac.uk/repository/xmlui/handle/20.500.12024/2539>
- Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 6(2), 159-177.
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication*, 33(2), 149-183.
- University of Warwick. (2025). BAWE (British Academic Written English) and BAWE Plus collections. Retrieved May 7, 2026, from <https://warwick.ac.uk/fac/soc/al/research/collections/bawe/>

THE AUTHORS

Bo Min Kim is a Visiting Research Fellow at the Institute of Foreign Language Education, Hankuk University of Foreign Studies, South Korea. She holds an MA in Teaching English to Speakers of Other Languages from the Institute of Education, University of London. She earned her PhD from Hankuk University of Foreign Studies, specializing in English language teaching methodology. She is the co-author of *A Walk Through English with My Child: A Smart Parent's Guide to English Learning* (2021). Her research interests include corpus linguistics, genre analysis, and Korean EFL learners' motivation to write.

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

Bo Min Kim

Visiting Research Fellow
Institute of Foreign Language Education
Hankuk University of Foreign Studies
107 Imun-ro, Dongdaemun-gu, Seoul, 02450, SOUTH KOREA
E-mail: b.m.kim.marieanna@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: 23 July 2026

Accepted: 15 August 2026