영어어문교육 *제31권* 2호 2025년 여름
https://doi.org/10.35828/etak.2025.31.2.155

고등학교 영어 교과서와 ChatGPT 생성 텍스트의 어휘 및 구문 특성 비교: 코퍼스 기반 분석

이영희[*]
(국립목포대학교)
강준수[**]
(중앙대학교)

Lee, Younghee Cheri, & Kang, Junsoo. (2025). A Corpus-Based Comparison of Lexical and Syntactic Features in High School English Textbooks and ChatGPT-Generated Texts. *English Language & Literature Teaching, 31*(2), 155-172.

This study empirically examines the strengths and limitations of generative AI texts as supplementary educational resources by comparing the lexical and syntactic features of high school English textbooks with those of texts generated by ChatGPT. A parallel corpus comprising 47 passages from ten government-authorized textbooks and their AI-generated counterparts was compiled for analysis. Four linguistic metrics were analyzed: number of word types, standardized type-token ratio (STTR), mean word length (MWL), and mean sentence length (MSL). WordSmith Tools 7.0 was used for corpus analysis, and statistical significance was assessed through independent t-tests. The results showed that while textbook passages contained a greater number of word types, ChatGPT texts demonstrated higher lexical diversity and longer mean word length. No significant difference was observed in sentence length. These findings suggest that AI-generated texts can offer enriched lexical input while maintaining structural clarity, thus serving as effective supplementary materials in classroom instruction. The study offers both theoretical insights and practical implications for integrating generative AI into secondary English education, especially for enhancing vocabulary exposure and expanding learner-centered reading resources.

[high school English textbook/ChatGPT/generative AI/lexical diversity/corpus, 고등학교 영어 교과서/ChatGPT/생성형 인공지능/어휘 다양성/코퍼스]

---

[*] 제1저자 [**] 교신저자

I. 서론

국가 교육과정의 변천과 함께 영어 교과서도 여러 차례 개정 과정을 거쳐 왔지만 여전히 EFL(English as a Foreign Language) 환경에서 교과서는 학습자의 주요 언어 입력원으로서 매우 중요한 위치를 차지한다(Vitta, 2023). 특히 수업 시간에 노출되는 교과서 텍스트는 학습자들에게 사실상 유일하거나 핵심적인 영어 입력이 되는 경우가 많으며, 이때 제공되는 어휘와 구문은 학습자의 이해도, 인지적 부담, 학습 동기 형성 등에 직접적인 영향을 미친다(Loi 외, 2022). 입력의 질적 적정성은 Krashen(1992)의 입력 가설에서도 강조되며, 교재는 학습자 수준보다 약간 도전적이지만 과도하게 어렵지 않은 형태로 구성되어야 한다. 실제로 어휘 및 문법 지식이 학습자의 읽기 및 쓰기 수행에 상이한 방식으로 영향을 미친다는 점이 보고된 바 있으며, 이는 교재 입력의 언어적 구성 요소를 고려한 설계의 중요성을 시사한다(배태일, 2025). 이러한 맥락에서 교과서에 포함된 어휘 난이도, 어휘 다양성, 문장 구조 복잡성 등을 정기적으로 점검하는 일은 영어교육에서 중요한 과제로 간주된다(강준수, 이영희, 2023a).

이러한 점검 작업은 전통적으로 코퍼스 언어학적 방법을 통해 이루어져 왔다. 교과서 간 어휘 난이도나 문장 길이를 비교하거나, 교과서와 원어민 코퍼스를 비교하여 실제 사용 빈도나 언어 수준 간의 괴리를 분석하는 연구들이 대표적이다(강준수, 이영희, 2023b). 이들 연구는 어휘 커버리지, 문장 가독성, 어휘 다양성(STTR) 등의 지표를 제시하고, EFL 학습자가 적절히 접해야 할 언어적 입력의 수준을 제안해 왔다(김지혜, 2023; 나혜성, 2024). 그러나 기존 연구 대부분은 교과서 자체에 초점을 맞추고 있어, 최근 교육적 관심이 높아지고 있는 생성형 AI 기반 텍스트와 교과서를 직접 비교한 실증 연구는 상대적으로 드문 실정이다(강준수, 이영희, 2023b).

2022년 말 출시된 ChatGPT를 비롯한 초거대 언어 모델은, 주어진 프롬프트에 따라 주제나 난이도를 조절해 자연스러운 영어 문장을 재구성할 수 있는 기능을 갖추고 있으며, 영어 교육 현장에서도 새로운 가능성을 제시하고 있다(강준수, 박신향, 2024). 특히 ChatGPT는 '쉬운 수준으로 요약해 달라'와 같은 요청에 따라 어휘 난이도와 문장 길이를 자동 조절하고, 어휘적 다양성과 문장 명료성 면에서도 우수한 결과를 산출한다는 보고가 있다(Botarleanu 외, 2020). 이러한 특성은 학습자에게 다양한 언어 자원을 제공하고, 교사에게는 학습자의 수준에 따라 텍스트 난이도를 조절할 수 있는 도구로 활용될 수 있다. 그러나 이와 같은 생성형 AI 텍스트가 실제 교과서와 비교하여 어휘 풍부함과 구문 복잡도 면에서 학습자에게 유의미한 입력인지에 대한 실증적 검토는 여전히 부족하다.

이에 본 연구는 고등학교 1학년 영어 교과서 원문과 ChatGPT가 동일 주제로 생성한 텍스트를 쌍 코퍼스로 구축하고, 양측 텍스트의 어휘 및 구문 특성을 정량 분석하여 AI 기반 언어 생성물이 학습 자료로서 가질 수 있는 가능성과 한계를 검토하고자 한다. 구체적인 연구문제는 다음과 같다.

1. 고등학교 영어 교과서 원문과 ChatGPT가 생성한 텍스트 간에 어휘 유형 수 및 어휘 다양성(STTR)은 어떠한 차이를 보이는가?
2. 고등학교 영어 교과서 원문과 ChatGPT가 생성한 텍스트 간에 평균 단어 길이와 평균 문장 길이는 어떠한 차이를 보이는가?

본 연구는 이 두 가지 연구문제에 대한 실증적 해답을 통해 교과서가 지닌 전통적 강점과 생성형 AI 텍스트가 제공할 수 있는 보완 가능성을 함께 조망하고자 한다. 특히 어휘 풍부함과 문장 복잡성 측면에서의 비교 · 분석을 바탕으로, EFL 환경에서 AI 기반 텍스트가 실제 수업에 어떻게 적용될 수 있을지 탐색하는 데 목적이 있다.

이러한 분석을 통해 확인된 차이와 공통점은 어휘 반복 노출, 문장 구조 난이도 조절, 학습자 맞춤형 자료 활용 등의 측면에서 교수 설계에 의미 있는 시사점을 제공할 것이다. 아울러 생성형 AI 텍스트가 창의적 표현이나 어휘 활용에서 교과서를 보완할 수 있는지, 또는 학습 부담을 유발할 가능성은 없는지 실증적으로 평가함으로써 교사와 교재 개발자에게 균형 잡힌 교육적 판단의 근거 자료를 제공하고자 한다.

결과적으로 본 연구는 교과서와 생성형 AI 텍스트 간의 상호보완 가능성을 실증적으로 제시함으로써, 학습자의 언어 입력 환경을 풍부하게 하고, 난이도 조절과 학습 효율성을 높이는 데 기여할 수 있는 전략을 제안하고자 한다. 또한 향후 다양한 수준과 맥락에서 생성형 AI의 효과적인 도입을 위한 교수 · 학습 설계에 유용한 기초 자료가 되기를 기대한다.

Ⅱ. 이론적 배경

1. 영어 교육에서 학습용 텍스트의 언어적 특성

영어 교육에서 학습용 텍스트의 언어적 특성은 학습자의 이해도와 인지 부하, 나아가 학습 성과에 중요한 영향을 미친다. Krashen(1992) 입력 가설에 따르면, 학습자는 현재 수준보다 약간 높은 난이도의 이해 가능한 입력(comprehensible input)을 제공받을 때 언어 습득이 촉진된다. 이러한 관점에서 교육용 텍스트는 어휘 및 구문 수준에서 학습자의 언어 수준과 인지적 특성을 고려하여 정교하게 설계될 필요가 있다.

특히 EFL 환경에서는 교과서와 같은 정제된 입력(refined input)이 학습자의 주요 언어 자원으로 기능하며, 텍스트에 포함된 어휘의 선택과 문장 구조는 학습자의 인지적 부담과 이해 가능성에 직접적인 영향을 준다. 학습자가 지속적으로 이해 가능한 입력을 접하기 위해서는 어휘 수준에서의 조절이 중요하다. 어휘 커버리지 연구에 따르면, 텍스트 이해를 위해서는 약 95~98%의 어휘를 학

습자가 인지하고 있어야 한다(Hsueh-Chao 외, 2000; Nation & Beglar, 2007). 이에 따라 교육용 텍스트는 고빈도 어휘 중심으로 구성되는 경우가 많으며, 어휘의 반복 노출을 통해 장기 기억을 유도하는 전략적 구성이 필요하다(Pellicer-Sánchez & Schmitt, 2010; Webb, 2007).

그러나 어휘의 다양성이 지나치게 제한될 경우 어휘 확장 기회가 줄어들 수 있기 때문에, 고빈도 어휘와 새로운 어휘 간의 균형 잡힌 배치가 중요하다. 구문적 측면에서도 문장의 복잡성은 학습자의 작업 기억 용량과 밀접하게 연관되어 언어 처리에 영향을 미친다. 학습 초기에는 단순하고 명료한 문장이 효과적이며, 점차 문장 구조의 복잡성을 조정하는 접근이 바람직하다. Flesch-Kincaid 등 읽기 용이성 지표는 이러한 난이도 조절을 위한 유용한 도구로 활용된다. 이전 연구들은 정제된 문장이 원문보다 학습자의 이해도 면에서 효과적임을 보여주며(Long & Ross, 1993; Yano 외, 1994), 이는 학습자 수준에 적합한 문장 구조 설계의 필요성을 뒷받침한다.

결국, 학습용 텍스트의 어휘적 · 구문적 특성은 학습자의 인지 부하를 조절하고 효과적인 언어 입력을 제공하기 위한 핵심 설계 요소이다(Paas 외, 2003). 특히 교과서 텍스트는 정제된 어휘 사용과 반복, 구조적 단순화를 통해 학습자가 과도한 인지 부담 없이 학습 내용을 내면화하도록 돕는다. 이처럼 텍스트의 언어적 특성은 단순한 내용 전달을 넘어 영어 교육에서 교수 설계의 정교화를 가능하게 하는 이론적 기반이 된다.

2. 영어 교과서 코퍼스 기반 분석 연구 동향

영어 교과서의 언어 입력을 코퍼스 언어학 방법론으로 분석하려는 시도는 수십 년간 활발히 진행되어 왔다. 대규모 텍스트 코퍼스를 기반으로 어휘 및 구문 지표를 계량적으로 분석함으로써, 교과서 간 어휘 분포, 난이도, 구조적 복잡성 등을 비교 · 평가한 연구들이 축적되어 있다(강준수, 박신향, 2024; 강준수, 이영희, 2024). 어휘 다양성, 어휘 수준, 평균 문장 길이, 평균 단어 길이, 가독성 지수 등의 지표는 교과서 간 언어 입력의 일관성과 적정 수준을 진단하는 데 활용되어 왔다.

국내 연구에서는 고등학교 영어 교과서 10종을 분석하여 교과서 간 어휘 및 구문 특성의 편차를 확인한 바 있다(강준수, 이영희, 2024). 이 연구는 동일한 교육과정을 기반으로 개발된 교과서들 간에도 어휘 수준, 어휘 다양성, 평균 어휘 길이, 문장 길이, 가독성 지수 등에서 통계적으로 유의미한 차이가 있음을 밝혔으며, 이는 언어 입력의 난이도나 구성 방식이 학습자에게 불균형한 학습 기회를 제공할 수 있음을 시사한다. 따라서 교육과정과의 정합성을 고려한 교과서 개발의 중요성이 강조된다.

한편, 중등 영어 및 불어 교과서를 대상으로 한 코퍼스 기반 연구에서는 어휘의 빈도 수준과 반복률이 어휘 학습 기회에 영향을 미칠 수 있음을 정량적으

로 분석하였다(Van Parys 외, 2024). 북한 영어 교과서를 코퍼스로 구축해 남북 교과서를 비교한 연구는, 서로 다른 교육 이념과 환경이 반영된 어휘 및 구문 사용의 차이를 보여준다(이영희, 김태영, 2022).

국외 연구에서도 교과서와 평가 지문, 또는 원어민 코퍼스 간의 비교를 통해 교과서 언어 입력의 한계를 진단하려는 시도가 이어졌다. 예컨대 터키 고등학교 영어 교과서와 대학 입시 지문 비교 연구에서는, 교과서의 어휘 수준과 문장 복잡성이 실제 시험 지문보다 낮다는 결과가 제시되었다(Gedik & Kolsal, 2022). Li(2022)는 중국인 영어 학습자의 조동사 사용 발달을 코퍼스 기반의 분석을 통해, 학습자의 언어 사용 양상을 이해하기 위해서는 실제 학습 입력 자료의 특성에 대한 분석이 선행되어야 함을 강조하였다.

최근에는 생성형 인공지능이 생성한 텍스트를 코퍼스 분석의 대상으로 포함하는 시도도 증가하고 있다. ChatGPT가 생성한 영어 텍스트의 어휘 및 구문 특성을 분석한 연구에 따르면, 해당 텍스트는 교과서와 유사한 구조를 가지면서도 어휘 다양성 등의 측면에서는 차별적인 특성을 보이는 것으로 나타났다(강준수, 박신향, 2024). 이는 AI 기반 언어 생성물이 학습용 텍스트로서 어떤 가능성과 한계를 갖는지 파악하는 데 코퍼스 분석이 유용한 방법론이 될 수 있음을 시사한다.

3. 생성형 AI 기반 영어 텍스트의 교육적 활용 가능성

생성형 AI의 등장은 영어교육 현장에 새로운 가능성을 열고 있다. 특히 2022년 말 출시된 OpenAI의 ChatGPT는 다양한 주제에 대해 자연스러운 문장을 생성할 수 있는 초거대 언어 모델로, 영어 작문, 회화 연습, 피드백 제공 등 학습 도구로 활용되는 사례가 급증하고 있다. 최근 실험 연구들을 종합한 리뷰에 따르면, ChatGPT는 영어 학습자의 문법 교정, 어휘 학습, 쓰기 능력 향상에 긍정적인 영향을 미치는 것으로 나타났다(Guo 외, 2024).

이러한 가능성은 생성형 AI 텍스트가 지닌 언어적 특성에 기초한다. Reviriego 외(2024)는 ChatGPT 초기 모델이 어휘 다양성과 표현력 측면에서 인간 작성 문장과 차이를 보였으나, 최신 모델은 어휘 다양성 측면에서 인간 수준에 근접하거나 능가하는 경향도 있다고 지적하였다. 그러나 일부 연구에서는 반복적이고 정형화된 표현 사용이 여전히 존재하며, 이는 학습자에게 다양한 표현 노출의 중요성을 시사한다(Kobak 외, 2024).

또한 ChatGPT는 정보 요약과 재구성 기능이 뛰어나며, 이는 복잡한 텍스트를 핵심 중심으로 압축하거나 쉬운 표현으로 바꾸는 등 학습자의 인지 부담을 줄이는 데 효과적이다. 요약 과정은 자기주도 학습을 가능하게 하지만, 정보 누락이나 왜곡 가능성이 있어 교사의 확인이 필요하다(Simelane 외, 2025).

문장의 명료성과 일관성도 강점으로 꼽힌다. 문법적으로 정확하고 명료한 문장을 생성하여 학습자의 이해를 돕고(Zakarneh 외, 2025), 특히 학습 초기 단계에서 정제된 입력으로 언어 구조 습득에 효과적이다. 다만 지나치게 일관된 문

체에만 노출될 경우 다양한 스타일 경험이 제한될 수 있으므로, 인간 텍스트와의 병행 사용이 권장된다.

무엇보다 ChatGPT는 난이도 조절과 개인화가 가능하다는 점에서 교과서와 차별화된다(Xu 외, 2024). 동일한 정보를 초급, 중급, 고급 수준으로 변환할 수 있으며, 프롬프트를 통해 어휘 수준과 문장 구조가 자동 조정된다. 실제로 ChatGPT는 프롬프트에 따라 문장 길이, 평균 단어 길이 등을 조절하고, 표현 다양성도 일정 수준 유지할 수 있다는 결과가 보고되었다(강준수, 이영희, 2024).

이처럼 교육적 활용의 장점이 존재하지만, 한계도 분명하다. 생성된 텍스트의 사실 여부가 불분명하거나, 문화적 맥락과 부합하지 않는 내용이 포함될 수 있어, 교사의 지도와 점검이 반드시 병행되어야 한다(강준수, 박신향, 2024). ChatGPT가 교과서 텍스트를 재구성한 결과를 분석한 연구에서도, 적절한 난이도 조절과 어휘 다양성은 가능하지만, 교육 자료로서의 활용에는 신중한 접근이 필요함이 제시되었다.

결론적으로 생성형 AI 기반 텍스트는 어휘 다양성, 정보 압축력, 문장 명료성, 난이도 조절력 등의 측면에서 영어 학습자에게 풍부하고 유연한 언어 입력을 제공할 수 있다(Evmenova 외, 2024). 이는 특히 언어 노출이 제한된 EFL 환경에서 효과적인 보완 자원으로 기능할 수 있으며, 전통 교과서와 생성형 AI 간의 유기적 결합 가능성을 시사한다.

Ⅲ. 연구방법

1. 코퍼스 데이터

본 연구에서 활용한 코퍼스는 고등학교 1학년 영어 공통 교과서의 읽기 지문 본문과, 이를 기반으로 ChatGPT를 통해 생성한 대응 텍스트로 구성되었다. 총 10종의 검정 교과서를 대상으로 하였으며, 각 교과서에서 4~6개의 단원을 선별하여 분석 대상으로 삼았다. 최종적으로 확보된 텍스트는 교과서 본문 47개와 이에 상응하는 GPT 생성 텍스트 47개로, 총 94개의 텍스트 파일로 구성되었다. 이와 같은 구성을 통해 동일 주제와 학습 목표를 지닌 텍스트 쌍 간의 정량적 · 정성적 비교가 가능하며, 분석의 일관성과 신뢰성 확보가 가능하다.

전체 어휘 수(Token)는 교과서 텍스트에서 38,547개, ChatGPT 생성 텍스트에서 31,797개로 나타났으며, 단원당 평균 어휘 수는 각각 820개와 676개였다. 표 1은 출판사별 텍스트 수, 단원 수, 어휘 수 등 주요 외형적 특징을 요약한 것이다. GPT 생성 텍스트에서 상대적으로 적은 어휘 수가 관찰된 것은 해당 모델이 텍스트를 재구성하는 과정에서 반복적이거나 불필요한 요소를 생략하거나 요약하는 특성이 반영된 결과이다.

이러한 어휘 수 차이는 GPT 생성 텍스트가 텍스트를 재구성하는 과정에서

반복 표현이나 불필요한 수식 요소 등을 최소화하여 간결하게 구성되었기 때문으로 해석할 수 있다(Dinçer 외, 2024). 실제로 GPT는 의미 전달에 핵심적인 정보 위주로 내용을 요약하며, 문장 간 중복을 줄이는 경향을 보인다. 이는 어휘 수 자체는 감소하였더라도, 정보의 밀도나 표현의 명료성 측면에서는 차별화된 특성을 보일 수 있음을 의미한다. 따라서 GPT 텍스트의 어휘 수 감소는 정보량의 단순 부족이라기보다는 생성 방식의 특성으로 이해할 수 있다.

코퍼스의 양은 상대적으로 제한적일 수 있으나, 본 연구는 고등학교 1학년 영어 공통 교과서를 출판사 전 범위에서 수집하여 구성했다는 점에서 높은 대표성을 가진다. 특히 국가 교육과정에 따라 검정된 교과서를 기반으로, 동일 주제의 내용을 ChatGPT가 재구성한 텍스트와 비교했다는 점에서, 생성형 AI 기반 언어 생성물과 교육용 정제 텍스트 간의 실질적 차이를 탐색하는 데 유의미한 기초자료로 기능한다.

ChatGPT 텍스트 생성을 위한 프롬프트는 교과서 각 본문의 주제와 학습 목표를 유지하면서, 고등학교 1학년 수준에 적합한 자연스럽고 명료한 영어 문장으로 재구성(paraphrasing)하도록 설계되었다. 실제로 사용한 기본 프롬프트는 다음과 같다.

> *"Paraphrase the following English text into more natural and authentic English while keeping the same meaning. Use clear structure and logical flow suitable for upper secondary students."*

이 프롬프트는 의미 보존을 전제로 하되, 문장 표현의 자연스러움, 논리적 구성, 학습자 수준 적합성을 강조한 것이 특징이다. 텍스트의 성격에 따라 일부 표현을 조정한 변형 프롬프트도 활용하였으며, 이에 대한 구체적인 예시는 부록에 수록하였다. 이와 같은 프롬프트 설계는 GPT 응답의 일관성과 재현 가능성을 확보하는 데 기여하였다.

표 1
교과서 vs. ChatGPT 코퍼스의 외형적 규모

출판사	교과서 어휘 수	GPT 어휘 수	단원 수
천재(강상구)	3,780	3,162	4
천재(조수경)	4,039	3,520	5
동아(이병민)	3,190	3,106	5
지학사(신상근)	3,979	3,149	4
미래엔(김성연)	3,392	2,677	4
능률(민병천)	3,815	3,355	6
능률(오선영)	4,245	3,196	5
비상(홍민표)	4,335	3,088	4
YBM(김은형)	4,143	3,716	6
YBM(박준언)	3,629	2,828	4
합계	38,547	31,797	47
평균(단원 기준)	820	676	

2. 연구 도구 및 분석 방법

본 연구는 고등학교 영어 교과서 원문과 ChatGPT가 재구성한 텍스트 간 어휘 및 구문적 특성의 차이를 정량적으로 분석하기 위해 코퍼스 기반 분석 방법을 채택하였으며, 텍스트 분석 도구로는 WordSmith Tools 7.0을 활용하였다. 본 연구의 구체적 분석 절차와 지표는 다음과 같다.

먼저, 어휘적 특성 분석은 텍스트의 어휘 구성 양상과 다양성을 정량적으로 파악하기 위해 어휘 유형 수(type)와 어휘 다양성(STTR: Standardized Type-Token Ratio)을 중심으로 이루어졌다. 어휘 유형 수는 텍스트 내 고유한 어휘의 총 개수로, 텍스트의 어휘적 풍부함을 보여주는 기초 지표이다. 반면 STTR은 동일 길이의 텍스트 간 어휘 분산과 반복성을 비교할 수 있는 대표적인 지표로, 어휘 사용의 다양성과 전략적 분포를 파악하는 데 유용하다(Nation, 2013). STTR은 일정 단위로 나뉜 텍스트 블록 내에서의 어휘 다양성을 반복적으로 계산한 뒤 평균값을 도출함으로써, 길이가 다른 텍스트 간 비교의 공정성을 확보하는 데 중요한 역할을 한다(강준수, 이영희, 2024). 본 연구에서는 텍스트 유형 간 어휘 사용의 전략성과 표현 범위의 차이를 보다 정밀하게 탐색하기 위해 STTR 지표를 주요 분석 지표로 설정하였다.

본 연구는 개별 어휘나 문장에 대한 미시적 접근보다, 텍스트 전반의 어휘적 특성을 파악하는 데 주안점을 두었다. 이는 생성형 인공지능 텍스트와 전통 교과서 텍스트 간의 언어적 차이를 객관화하여 비교 가능한 틀을 제공하고, 텍스트 유형 간 언어 구성의 전략적 차이를 실증적으로 탐색할 수 있다는 점에서 학문적 의의를 가진다. 특히 STTR, 평균 단어 길이, 어휘 유형 수와 같은 핵심 지표는 코퍼스 기반 언어 비교 연구에서 언어 다양성 및 복잡성의 정량적 평가를 가능하게 하는 필수 지표로, 학습 자료의 언어적 품질을 진단하는 기초 자료로 기능할 수 있다.

다음으로, 구문적 특성 분석에서는 평균 단어 길이(MWL: Mean Word Length)와 평균 문장 길이(MSL: Mean Sentence Length)를 측정하였다. 평균 단어 길이는 텍스트에 포함된 단어들의 평균 문자 수를 나타내며, 어휘의 추상화 수준과 복잡도를 반영하는 지표로 학습자의 인지적 부담과도 직결된다. 평균 문장 길이는 한 문장이 포함하는 단어 수의 평균으로, 문장의 구조적 복잡성과 이해 용이성을 판단하는 핵심 요소로 사용된다. 이 두 지표는 선행연구들에서도 교과서 텍스트의 일관성, 난이도 조절 여부 등을 평가할 때 활용된 대표적 기준이다(강준수, 이영희, 2023a).

이상의 네 가지 지표에 대해 각각 독립표본 *t*-검정을 실시하여, 교과서 텍스트와 ChatGPT 텍스트 간에 통계적으로 유의미한 차이가 존재하는지를 확인하였다. 이러한 분석틀은 두 텍스트 유형 간 언어적 특성 차이를 명확히 도출할 뿐 아니라, ChatGPT 기반 생성 텍스트의 교육적 활용 가능성과 교육 자료로서의 적합성을 판단하는 실질적인 기준으로 기능할 수 있다는 점에서 연구적 의

의가 있다.

Ⅳ. 연구결과 및 논의

본 연구는 고등학교 영어 교과서 본문과 ChatGPT가 재구성한 텍스트를 대상으로 어휘적 및 구문적 특성을 비교·분석함으로써, 생성형 인공지능 언어 생성 도구의 교육적 활용 가능성과 한계를 실증적으로 탐색하고자 하였다. 이를 위해 어휘 유형 수, 어휘 다양성(STTR), 평균 단어 길이, 평균 문장 길이를 주요 분석 지표로 설정하고, 각각에 대해 *t*-검정을 실시하여 두 텍스트 유형 간의 통계적 차이를 확인하였다. 본 장에서는 분석 결과를 어휘적 특성과 구문적 특성의 두 범주로 나누어 제시하고, 각 결과가 갖는 교육적 시사점을 함께 논의하고자 한다.

1. 어휘적 특성 분석 결과

첫 번째 연구문제에 대한 해답을 도출하기 위해, 고등학교 영어 교과서 본문과 ChatGPT가 재구성한 텍스트 간 어휘적 특성을 분석하였다. 어휘적 특성은 어휘 유형 수(type)와 어휘 다양성을 중심으로 분석하였으며, 독립표본 *t*-검정을 실시하였다(표 2 및 그림 1 참조).

분석 결과, 교과서 텍스트의 어휘 유형 수는 평균 363.640(SD = 78.227)으로, ChatGPT 텍스트의 평균 341.360(SD = 62.624)보다 유의미하게 높게 나타났다($t(46) = 3.673$, $p < .001$). 반면, 어휘 다양성은 ChatGPT 텍스트가 평균 97.633(SD = 0.601)으로, 교과서 텍스트의 평균 96.923(SD = 0.795)보다 유의하게 높았다($t(46) = -5.749$, $p < .001$).

이 결과는 전체 어휘 수(token 수) 기준으로는 ChatGPT 텍스트가 교과서 텍스트보다 적었으나, 어휘 다양성 지표(STTR)와 어휘 유형 수(type 수) 분석 결과를 종합적으로 고려하면, 반복 어휘의 빈도는 낮고 다양한 어휘가 보다 고르게 사용되는 경향이 관찰되었다. 즉, ChatGPT 텍스트의 STTR 값이 더 높다는 것은 어휘 사용의 다양성이 상대적으로 뛰어남을 의미한다. 본 연구는 이와 같은 어휘 사용 경향을 보다 명시적으로 보여주기 위해, 각 텍스트의 고빈도 어휘 목록과 빈도 분포를 부록에 수록하였다(부록 참조). 이는 학습자에게 신선하고 확장된 어휘 자극을 제공할 수 있다는 점에서 교육적으로 긍정적인 특성을 가진다.

다만, STTR 지표는 텍스트의 길이에 일정 부분 민감하게 반응하는 특성이 있으므로, ChatGPT 텍스트의 상대적으로 짧은 평균 어휘 수(676개)가 STTR 값의 상승에 영향을 주었을 가능성도 존재한다. 본 연구는 가능한 한 동일한 주제와 길이 조건을 유지하고자 하였으나, 해석 시 이러한 지표 특성을 함께 고려할

필요가 있다.

그러나 동시에, 어휘 다양성이 높다는 점은 반복 노출을 통한 어휘 정착 기회가 줄어들 수 있음을 의미하기도 한다. 특히 EFL 환경에서는 어휘의 반복적 노출이 장기 기억에 긍정적 영향을 미치는 중요한 학습 전략으로 간주되기 때문에, ChatGPT 텍스트의 높은 다양성이 인지적 부담 요인이 될 수 있다는 우려도 함께 제기된다(김태영, 이영희, 2022; 한지수, 이영희, 2022).

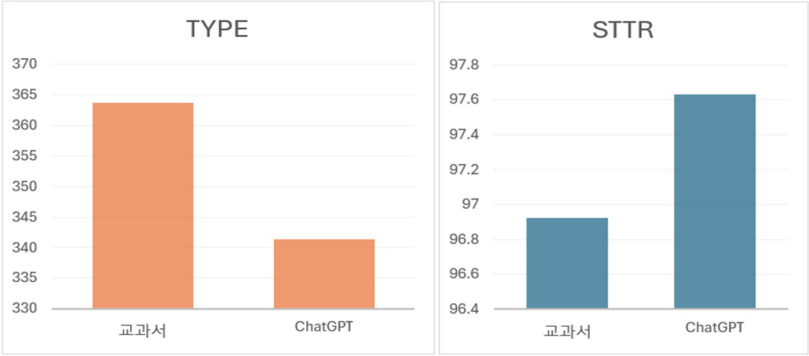
따라서 본 분석 결과는 교과서 텍스트가 절대적인 어휘량에서 우위를 가지며 기초 어휘의 반복 학습에 적합한 반면, ChatGPT 텍스트는 확장적 어휘 노출 및 다양성 제고에 유리하다는 상호보완적 특성을 시사한다. 학습자의 수준, 학습 목표, 수업 맥락에 따라 두 자료의 병행 활용이 이루어진다면, 보다 균형 잡힌 어휘 학습 환경을 설계할 수 있을 것이다.

표 2
어휘 유형 수(TYPE)와 어휘 다양성(STTR)의 비교 결과

지표	텍스트 유형	*M*	*SD*	*t*(46)	*p*
TYPE	교과서	363.640	78.227	3.673	.001***
	ChatGPT	341.360	62.624		
STTR	교과서	96.923	0.795	-5.749	.001***
	ChatGPT	97.633	0.601		

* $p < .05$, ** $p < .01$, *** $p < .001$

그림 1
교과서 vs. ChatGPT 텍스트 간 어휘 TYPE 및 STTR의 비교



2. 구문적 특성 분석 결과

두 번째 연구문제에 대한 해답을 도출하기 위해, 고등학교 영어 교과서 본문과 ChatGPT가 재구성한 텍스트 간의 구문적 특성 차이를 분석하였다. 구문적 특성은 평균 단어 길이(MWL: Mean Word Length)와 평균 문장 길이(MSL: Mean Sentence Length)를 중심으로 검토되었으며, 독립표본 *t*-검정을 실시하였다(표 3 및 그림 2 참조).

분석 결과, 평균 단어 길이에 있어서는 ChatGPT 텍스트의 평균값이 4.682(*SD* = 0.243)로, 교과서 텍스트의 평균 4.632(*SD* = 0.306)보다 유의미하게 더 긴 것으로 나타났다*(t*(46) = -2.064, *p* = .045). 이는 ChatGPT가 단어 선택에 있어 보다 복합적이고 길이가 긴 어휘를 활용하는 경향이 있음을 시사하며, 그 결과 학습자에게 더 높은 인지적 부담을 유발할 가능성도 있다. 이러한 경향은 어휘의 길이 차이가 학습자의 텍스트 이해도와 난이도 인식에 영향을 줄 수 있다는 기존 연구 결과와도 일치한다(강준수, 이영희, 2024).

평균 어휘 길이는 ChatGPT 텍스트가 4.682, 교과서 텍스트가 4.632로 나타나 통계적으로 유의미한 차이를 보였으나, 수치 차이는 크지 않았다. 단어 길이 분포를 살펴본 결과, 두 텍스트 모두 가장 긴 단어는 각각 16자(예: *responsibilities*, *environmentally*)였으며, 1자 단어(a, I)의 사용 빈도는 ChatGPT 텍스트(1,543개)가 교과서 텍스트(1,292개)보다 더 많았다. 이는 ChatGPT 텍스트에서 상대적으로 관사 ‘a’나 대명사 ‘I’와 같은 기능어(function words)의 사용 빈도가 높았음을 시사하며, 결과적으로 평균 어휘 길이 증가에 기여했을 가능성이 있다. 따라서 평균 어휘 길이의 통계적 유의미성은 어휘 구성의 구조적 특성과 문장 구성 방식의 차이에서 비롯된 것으로 해석할 수 있다.

반면, 평균 문장 길이는 교과서 텍스트가 평균 15.589(*SD* = 2.719), ChatGPT 텍스트가 평균 15.023(*SD* = 2.399)으로 나타났으며, 두 텍스트 간 통계적으로 유의미한 차이는 관찰되지 않았다(*t*(46) = 1.909, *p* = .063). 이는 두 텍스트 간 문장의 구조적 복잡도나 길이가 전반적으로 유사한 수준임을 보여주는 결과이며, ChatGPT가 사용자의 특별한 요청 없이도 교과서 수준의 문장 구성 명료성을 재현할 수 있다는 점을 시사한다.

실제 선행연구에서도, ChatGPT는 문장 길이를 유지하면서도 문장 내 정보 구조화나 문체 단순화를 통해 읽기 용이성을 향상시킬 수 있음이 확인된 바 있다(강준수, 이영희, 2024). 이러한 결과는 ChatGPT 생성 텍스트가 문장 단위의 구조적 일관성과 명료성을 갖추고 있음을 보여주며, 학습자의 수준에 맞는 보조 학습 자료로 활용될 가능성을 충분히 내포하고 있다.

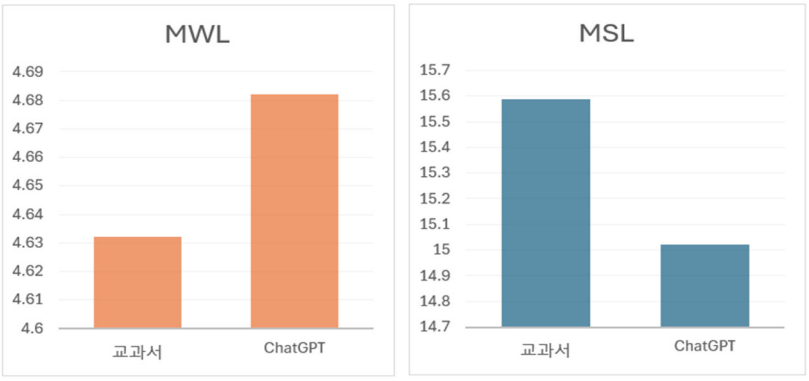
다만, 평균 문장 길이만으로는 문장의 실제 인지적 복잡성을 완전히 설명하기에는 한계가 있으며, 향후에는 Gunning Fog Index나 Flesch-Kincaid Grade Level과 같은 읽기 이독성(readability) 지표를 병행하여 구문 난이도를 보다 정교하게 해석할 수 있을 것이다.

표 3
평균 단어 길이(MWL)와 평균 문장 길이(MSL)의 비교 결과

지표	텍스트 유형	*M*	*SD*	*t*(46)	*p*
MWL	교과서	4.632	0.306	-2.064	045*
	ChatGPT	4.682	0.243		
MSL	교과서	15.589	2.719	1.909	063
	ChatGPT	15.023	2.399		

* $p < .05$, ** $p < .01$, *** $p < .001$

그림 2
교과서 vs. ChatGPT 텍스트 간 MWL 및 MSL의 비교



이상과 같이, 본 연구는 고등학교 영어 교과서 본문과 ChatGPT가 생성한 대응 텍스트를 대상으로 어휘 및 구문 수준에서의 언어적 특성을 실증적으로 비교 · 분석하였다. 분석 결과, 어휘적 측면에서는 ChatGPT 텍스트가 더 높은 어휘 다양성(STTR)을 보인 반면, 절대 어휘 수(TYPE)에서는 교과서가 더 풍부한 어휘량을 제공하는 것으로 나타났다. 구문적 측면에서는 ChatGPT 텍스트가 평균 단어 길이에서 다소 더 복잡한 양상을 보였으나, 문장 길이는 교과서와 유사한 수준으로 유지되었다.

이러한 결과는 ChatGPT 기반 생성 텍스트가 어휘 다양성과 문장 명료성 측면에서 교육적 활용 가능성을 지니며, 특히 교과서와 병행 활용될 경우 학습자 수준과 목표에 맞춘 균형 잡힌 언어 입력 자료로 기능할 수 있음을 시사한다. 다음 장에서는 본 연구 결과를 바탕으로 한 교육적 시사점과 함께, 향후 연구를 위한 제언을 제시하고자 한다.

V. 결론 및 제언

EFL 환경에서 교과서는 학습자의 핵심 언어 입력 자료로 기능하며, 그 언어적 특성은 학습자의 이해도, 인지 부하, 학습 성과에 직접적인 영향을 미친다.

그러나 교과서의 언어 입력이 학습자에게 얼마나 효과적이고 적절한지를 실증적으로 점검하려는 시도는 여전히 부족한 실정이다. 특히 최근에는 ChatGPT와 같은 생성형 인공지능이 교육 현장에 활용되기 시작했지만, 해당 텍스트가 전통적인 교과서와 비교하여 언어적 타당성을 확보하고 있는지에 대한 검토는 미흡하다. 이에 본 연구는 고등학교 영어 교과서와 ChatGPT가 생성한 텍스트를 어휘 및 구문적 측면에서 정량적으로 비교함으로써, AI 기반 텍스트의 교육적 가능성과 한계를 실증적으로 진단하고자 하였다.

연구 결과, 어휘적 측면에서는 ChatGPT 텍스트가 더 높은 어휘 다양성을 보이는 반면, 절대 어휘 수에서는 교과서가 우위를 나타냈다. 이는 ChatGPT가 반복을 줄이고 다양한 어휘를 전략적으로 배치하는 경향을 보임을 의미하며, 디지털 기반 어휘 교육에서 활용 가치가 있음을 시사한다. 특히 어휘 자극의 확장성이 필요한 학습 상황에서 보조 자료로 효과적으로 기능할 수 있다. 이러한 결과는 교과서 텍스트가 절대적인 어휘량에서 우위를 가지며 기초 어휘의 반복 학습에 적합한 반면, ChatGPT 텍스트는 확장적 어휘 노출 및 다양성 제고에 유리하다는 상호보완적 특성을 시사한다.

더불어, ChatGPT 텍스트의 높은 어휘 다양성과 평균 단어 길이는 일부 학습자에게는 어휘 확장의 기회로 인식될 수 있으나, 다른 학습자에게는 다소 높은 난이도로 인식되어 인지적 부담을 줄 가능성도 있다.

이처럼 언어적 특성은 학습자의 수준 및 인식과 상호작용할 수 있으므로, 실제 수업 적용 시에는 학습자의 반응을 고려한 텍스트 선택과 교수설계가 병행되어야 할 것이다.

구문적 측면에서는, ChatGPT 텍스트가 평균 단어 길이에서 교과서보다 다소 높은 수치를 보이며 더 복잡한 어휘를 사용했지만, 평균 문장 길이에서는 교과서와 유사한 수준을 유지하였다. 이는 AI 텍스트가 복합 어휘를 포함하더라도 문장 수준에서는 학습자의 인지적 부담을 고려한 구조적 명료성을 갖출 수 있음을 보여준다. 이와 같은 특성은 난이도 조절이 필요한 교수설계에 적합한 자료로서 AI 텍스트의 가능성을 뒷받침한다.

이러한 결과는 ChatGPT 기반 생성 텍스트가 어휘 다양성과 구조적 명료성을 바탕으로 학습 자료로서의 교육적 활용 가능성을 내포하고 있음을 시사한다. 특히 교과서와의 병행 활용을 통해 학습자의 수준과 목표에 맞춘 균형 잡힌 언어 입력 설계가 가능하다는 점에서 그 의의가 크다.

교과서와 ChatGPT 텍스트를 병행하여 제시하는 방식은 학습자에게 다양한 언어적 입력을 제공함으로써 어휘 선택, 문장 구성, 논리 전개 방식의 차이를 비교 분석하는 활동으로 확장될 수 있다. 이러한 활동은 학습자의 비판적 읽기 능력과 메타인지적 사고를 촉진하며, 생성형 AI 텍스트의 구조적 명료성과 어휘 다양성을 활용한 학습 자료 구성은 교사의 교수 설계 유연성을 높이는 데에도 기여할 수 있다.

그러나 본 연구는 다음과 같은 한계점도 지닌다. 첫째, 분석 대상이 고등학교

1학년 교과서와 일부 ChatGPT 응답에 한정되어 있어 결과의 일반화에는 제약이 있다. ChatGPT 응답은 주제, 문체, 프롬프트 방식에 따라 다양한 양상을 보일 수 있으므로, 향후 연구에서는 다양한 학년, 텍스트 유형, 프롬프트 조건을 포괄하는 코퍼스 기반 확장 연구가 필요하다.

둘째, 본 연구는 어휘 및 구문이라는 정량적 언어 특성에 국한되어 있으며, 담화 연결성, 문체 적합성, 학습자 반응성 등 질적 요소는 포함되지 않았다. 학습자가 AI 텍스트를 어떻게 이해하고, 얼마나 수용 가능한 자료로 인식하는지는 실험 설계와 질적 연구를 통해 보완되어야 한다. 특히 학습자 중심의 인터뷰, 인식 조사, 반응 기반 실험 연구는 AI의 교육적 타당성을 보다 풍부하게 조명할 수 있다.

그럼에도 불구하고, 본 연구는 생성형 AI가 제공하는 언어 입력이 교육 자료로서 어느 정도의 언어적 품질과 전략성을 확보하고 있는지를 실증적으로 검토하였다는 점에서 의의를 가진다. 어휘적 다양성과 구문적 명료성을 모두 갖춘 텍스트로서, ChatGPT는 읽기 자료 보완, 어휘 확장 활동, 문장 재구성 학습, 개별 수준 조정 등 다양한 교수전략에 활용될 수 있다.

더 나아가, ChatGPT 기반 텍스트는 단순 보조 자료를 넘어, 자기주도 학습의 동기 자극 요소로도 기능할 수 있다. 학습자가 자신의 수준에 맞는 텍스트를 직접 생성 · 수정하고 활용함으로써 개인화된 학습 환경을 구성하는 데 기여할 수 있으며, 교수자에게는 학습자의 반응에 따라 텍스트 난이도와 범위를 유동적으로 조절할 수 있는 새로운 가능성을 제공한다. 이러한 능동적 활용은 학습자의 기술 활용 역량에 따라 달라질 수 있으며, 특히 AI 기반 자료의 비판적 수용과 활용 능력은 임정완(2024)의 연구에서 논의된 디지털 환경 적응 역량과도 맞닿아 있다.

결론적으로, 본 연구는 생성형 인공지능 기반 텍스트가 전통적인 교과서와 비교하여 어느 수준의 언어적 특성을 지니고 있으며, 실제 교육 현장에서 어떻게 활용될 수 있을지를 실증적으로 조명하였다. 교과서와의 비교를 통해 AI 텍스트의 장단점을 정량적으로 분석하고, 교육적 적용 가능성을 구조화함으로써 향후 AI와 인간 중심 교육 자료의 공존 전략 수립에 기초 자료를 제공하였다는 점에서 그 의의가 크다. 특히, ChatGPT와 같은 생성형 AI는 교사가 제시한 특정 주제나 학습 수준에 맞추어 맞춤형 예문, 설명 문단, 배경 지식 보완 자료를 빠르게 생성할 수 있어, 어휘 · 문장 구조 중심의 글쓰기 수업, 독해 전략 훈련, 텍스트 간 비교 분석 활동 등 다양한 영어 교과 수업에서 보조 자료로 활용될 수 있다. 향후 연구는 이러한 분석을 기반으로 보다 다양한 교육 환경과 수준을 반영한 확장적, 통합적 접근을 시도함으로써 AI 기반 영어 교육의 정착을 위한 학문적 · 실천적 토대를 마련해 나갈 필요가 있겠다.

아울러, 생성형 AI 텍스트의 활용은 저작권, 문화적 편향, 사회적 수용성 등과 관련한 법적·윤리적 쟁점을 수반할 수 있으므로, 교육 현장 적용 시에는 이러한 요소에 대한 충분한 인식과 검토가 병행되어야 할 것이다.

참고문헌

강준수, 박신향. (2024). 영어 학습 보조도구로서 ChatGPT 연구: 어휘적 · 구문적 코퍼스 분석을 중심으로. *언어과학연구, 108*, 49-66. http://dx.doi.org/10.21296/jls.2024.03.108.49

강준수, 이영희. (2023a). 언어 입력의 일관성에 관한 코퍼스 기반 연구. *영어학연구, 29*(1), 127-151. https://doi.org/10.17960/ell.2023.29.1.007

강준수, 이영희. (2023b). 고등학교 영어교과서 코퍼스 기반 언어 입력의 일관성 탐구. *영어학연구, 29*(3), 161-194. https://doi.org/10.17960/ell.2023.29.3.008

강준수, 이영희. (2024). 디지털 리터러시 향상을 위한 ChatGPT 응답의 언어적 특성 분석: 코퍼스 연구. *영어학연구, 30*(2), 57-79. https://doi.org/10.17960/ell.2024.30.2.003

배태일. (2025). 영어 문법과 어휘가 영어 읽기 및 쓰기에 미치는 상대적인 영향 조사: 중·고등학생 비교. *영어어문교육, 31*(1), 59-76. https://doi.org/10.35828/etak.2025.31.1.59

이영희, 김태영. (2022). 북한 영어교육의 패러다임 전환: 코퍼스 기반 분석. *영어학, 22*(1), 279-299. https://doi.org/10.15738/kjell.22..202203.279

임정완. (2024). 중학생과 대학생 예비영어교사의 디지털 리터러시 역량 비교. *영어어문교육, 30*(3), 87-107. https://doi.org/10.35828/etak.2024.30.3.87

한지수, 이영희. (2022). 2022학년도 수능영어의 체감 난이도 상승 요인에 관한 코퍼스 기반 분석. *영어학, 22*(1), 336-354. https://doi.org/10.15738/kjell.22..202203.336

Dinçer, A. M., Axinte, S. D., Tod-Raileanu, G., & Bacivarov, I. C. (2024). AI tools introduced in software development: Analysis of code quality, security and productivity implications. In *2024 IEEE 30th International Symposium for Design and Technology in Electronic Packaging (SIITME)* (pp. 32-39). IEEE. https://doi.org/10.1109/SIITME59989.2024.10385459

Evmenova, A. S., Borup, J., & Shin, J. K. (2024). Harnessing the power of generative AI to support all learners. *TechTrends, 68*(4), 820-831. https://doi.org/10.1007/s11528-024-00966-x

Gedik, T. A., & Kolsal, Y. S. (2022). A corpus-based analysis of high school English textbooks and English university entrance exams in Turkey. *Theory and Practice of Second Language Acquisition, 8*(1), 157-176.

Guo, Q., Cao, J., Xie, X., Liu, S., Li, X., Chen, B., & Peng, X. (2024, February). Exploring the potential of chatgpt in automated code refinement: An empirical study. In *Proceedings of the 46th IEEE/ACM International Conference on Software*

Engineering (pp. 1-13). IEEE Computer Society. https://doi.org/10.1145/3597503.3623306

Hsueh-Chao, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430. https://doi.org/10.48550/arXiv.2406.07016

Kobak, D., González-Márquez, R., Horvát, E. Á., & Lause, J. (2024). Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv.* https://doi.org/10.48550/arXiv.2406.07016

Krashen, S. (1992). The input hypothesis: An update. *Linguistics and language pedagogy: The state of the art*, 409-431.

Li, L. X. (2022). Developmental patterns of English modal verbs in the writings of Chinese learners of English: a corpus-based approach. *Cogent Education, 9*(1), 1-23. https://doi.org/10.1080/2331186X.2022.2050457

Loi, N. V., & Thanh, D. T. K. (2022). Engaging EFL Learners in Reading: A Text-Driven Approach to Improve Reading Performance. *Tesl-Ej, 26*(2), 1-18. https://doi.org/10.55593/ej.26102a5

Long, M. & S. Ross (1993). Modifications that preserve language and content. In M. Tickoo (Ed.), *Simplification: Theory and application*. (pp. 29-52). SEAMEO Regional Language Centre.

Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*, 9-13.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational psychologist, 38*(1), 1-4. https://doi.org/10.1207/S15326985EP3801_1

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart?. *Reading in a Foreign Language, 22*(1), 31-55. http://hdl.handle.net/10125/66652

Reviriego, P., Conde, J., Merino-Gómez, E., Martínez, G., & Hernández, J. A. (2024). Playing with words: Comparing the vocabulary and lexical diversity of ChatGPT and humans. *Machine Learning with Applications, 18*, Article 100602. https://doi.org/10.1016/j.mlwa.2024.100602

Simelane, P. M., & Kittur, J. (2025). Use of Generative Artificial Intelligence in Teaching and Learning: Engineering Instructors' Perspectives. *Computer Applications in Engineering Education, 33*(1), Article e22813. https://doi.org/10.1002/cae.22813

Van Parys, A., De Wilde, V., Macken, L., & Perez, M. M. (2024). Vocabulary of reading materials in English and French L2 textbooks: A cross-lingual corpus study. *System*, 124, Article 103396. https://doi.org/10.1016/j.system.2024.103396

Vitta, J. P. (2023). The functions and features of ELT textbooks and textbook analysis: A concise review. *RELC Journal, 54*(3), 856-863. https://doi.org/10.1177/0033688221103582

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65. https://doi.org/10.1093/applin/aml048

Xu, X., Wang, X., Zhang, Y., & Zheng, R. (2024). Applying ChatGPT to tackle the side effects of personal learning environments from learner and learning perspective: An interview of experts in higher education. *PLoS One, 19*(1), Article e0295646. https://doi.org/10.1371/journal.pone.0295646

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language learning, 44*(2), 189-219.

Zakarneh, B., Annamalai, N., Al Said, N., & Aljabr, F. (2025). Revolutionizing language learning through ChatGPT: An analysis of English language learners. International *Journal of English Language and Literature Studies, 14*(1), 1-16. https://doi.org/10.55493/5019.v14i1.5274

부록

고빈도 어휘 목록 및 빈도 분포표

교과서					
순위	유형	빈도	순위	유형	빈도
1	people	161	26	get	39
2	like	98	27	called	38
3	make	76	28	feel	38
4	day	74	29	food	38
5	new	71	30	take	38
6	time	71	31	find	37
7	world	63	32	example	36
8	English	62	33	started	36
9	way	62	34	long	35
10	AI	57	35	something	35
11	years	55	36	boy	34
12	help	54	37	great	34
13	good	50	38	look	34
14	different	49	39	come	33
15	life	47	40	found	33
16	art	46	41	things	33

ChatGPT					
순위	유형	빈도	순위	유형	빈도
1	people	141	26	example	38
2	like	120	27	two	37
3	time	73	28	better	36
4	first	68	29	think	36
5	make	68	30	try	36
6	something	67	31	long	33
7	help	60	32	part	33
8	new	58	33	get	32
9	day	56	34	once	32
10	feel	52	35	coffee	31
11	AI	50	36	helps	31
12	made	50	37	see	31
13	world	48	38	today	30
14	different	45	39	used	30
15	English	44	40	called	29
16	way	44	41	real	29

17	coffee	46	42	better	32
18	school	45	43	place	32
19	two	45	44	said	32
20	old	44	45	information	31
21	see	44	46	know	31
22	important	43	47	reefs	31
23	used	43	48	very	31
24	think	40	49	work	31
25	use	40	50	friends	30

17	good	43	42	story	29
18	life	43	43	things	29
19	school	42	44	city	28
20	use	42	45	found	28
21	years	41	46	know	28
22	art	40	47	said	28
23	look	40	48	someone	28
24	take	39	49	waste	28
25	work	39	50	come	27

* ※ 본 표는 어휘적 특성 분석의 일환으로, 의미를 실질적으로 전달하는 내용어만을 대상으로 구성함. 기능어로 분류되는 대명사, 조동사, 전치사, 접속사, 한정사 등은 분석에서 제외하였음.

예시언어(Examples in): English
적용가능 언어(Applicable Languages): English
적용가능 수준(Applicable Levels): Secondary

이영희(교수)
국립목포대학교 글로벌커뮤니케이션학부 영어영문학전공
58554 전남 무안군 청계면 영산로 1666
Email: cheriberry@mnu.ac.kr

강준수(연구원)
중앙대학교 일반대학원 교육학과
06973 서울특별시 동작구 흑석로 84
Email: anf7788@cau.ac.kr

Received: April 14, 2025
Reviewed: June 10, 2025
Revised version received: June 25, 2025