CC BY

Asia Pacific Journal of Corpus Research
Vol. 6, No. 1

http://www.icr.or.kr/apjcr
e-ISSN 2733-8096
https://doi.org/10.22925/apjcr.2025.6.1.47

2022 개정 영어과 교육과정 실행을 위한 COCA(Corpus of Contemporary American English)의 교육적 활용 방안 연구

김태국
(인제고등학교)

Kim, T. G. (2025). Study on the educational applications of the Corpus of Contemporary American English (COCA) for the 2022 revised English curriculum. *Asia Pacific Journal of Corpus Research, 6*(1), 47-53.

The purpose of this study is to explore the educational applications of the Corpus of Contemporary American English (COCA) to enhance English communicative competence in alignment with the 2022 revised English curriculum in Korea. This paper analyzes COCA's primary functions—List, Chart, Word, Collocates, Compare, and KWIC—and proposes pedagogical methods for their use in the classroom. The study reveals three major findings. Firstly, COCA's extensive, genre-balanced data provides authentic language contexts for learning vocabulary and grammar. Secondly, its analytical tools enable learners to understand lexico-grammatical patterns, collocations, and genre-specific language variations, thereby improving their writing and reading skills. Finally, while COCA is a powerful resource, its limitations include a focus on American English and access restrictions for free users. Web-based corpora like COCA can foster a new paradigm of data-driven, self-directed learning. This approach effectively supports the goals of the 2022 curriculum by equipping learners with practical English skills and critical thinking abilities, making it a valuable supplementary tool for educators.

Keywords: Corpus of Contemporary American English (COCA), 2022 Revised English Curriculum, Data-Driven Learning (DDL), English Education, English Teaching and Learning

1. 서론

2022 영어과 개정교육과정은 영어 의사소통 역량을 기르는 것을 총괄 목표로 제시하고 있으며 교수·학습 및 평가의 방향을 디지털 인공지능 교육환경을 고려하여 구체적으로 제시하고 있다(교육부, 2022). 특히 인공지능 기술 및 디지털 도구로서 코퍼스 데이터를 활용하거나 코퍼스 활용 프로그램을 사용하여 영어 수행 능력을 평가하도록 제시하고 있다. 코퍼스는 대규모 언어 텍스트로 실제 언어 사용 맥락을 풍부하게 제시할 수 있다는 장점이 있다(권혁승, 정채관, 2012). 또한 정제된 교과서 텍스트만으로는 실제적이며 다양한 언어 사용 양상을 접하는 것은 어려운 실정이다. 따라서 실제 언어를 파악하고 교육 현장에서 사용하기 편리한 온라인 코퍼스와 프로그램이 필요하기에 10억 어절 규모와 균형 잡힌 장르 구성을 갖춘 Corpus of Contemporary American English(COCA)는 유용한 교육 자료라 할 수 있다. 본 논문은 COCA의 주요 기능과 특징을 분석하고, 이를 활용한 교수·학습 방안을 제시하여 2022 개정 교육과정이 요구하는 영어 의사소통 역량 함양을 지원하고자 한다.

2. COCA 주요 기능

COCA는 사용자가 방대한 텍스트 데이터 내에서 특정 언어 패턴을 효과적으로 발견하고 분석할 수 있도록 다양한 웹 기반 인터페이스와 분석 도구를 제공한다. 직관적인 메뉴를 통해 사용자는 복잡한 검색(query)을 실행하고 그 결과를 다각도로 시각화할 수 있다. 본 장에서는 COCA의 핵심 기능인 List, Chart, Word, Collocates, Compare, KWIC(Key Word in Context)를 중심으로 각각의 기능과 활용법을 구체적인 예시와 함께 제시하고자 한다. 각 기능을 사용하기 위해서는 등록부터 해야 하며 무료인 경우 하루에 20번 검색할 수 있다.

2.1. List 기능

List 기능은 COCA의 가장 기본적인 검색 도구로, 검색 단어의 어휘 빈도를 산출하는 기능을 제공한다. 뿐만 아니라 상단에 IMAGE, PRON/VIDEO, BOOK, THESAURUS 기능을 제공한다. 각 기능을 선택하고 검색된 단어를 클릭하면, 검색된 단어와 관련된 이미지, YouENGlish를 통한 영상 및 발음 연습 자료, 검색된 단어가 사용된 책 정보, 동의어 등을 확인할 수 있다. 이러한 기능은 교수·학습 자료로서 다양하게 사용될 수 있음을 확인할 수 있다.

단어를 검색할 때 다양한 방법으로 검색할 수 있다. 와일드 카드(literal*), 여러 단어 검색(see | watch), 레마 검색([talk]), 품사 검색(waterVERB)를 사용하여 다양한 어휘를 검색할 수 있다. 뿐만 아니라 구(phrase)도 비슷한 방식으로 검색할 수 있다.

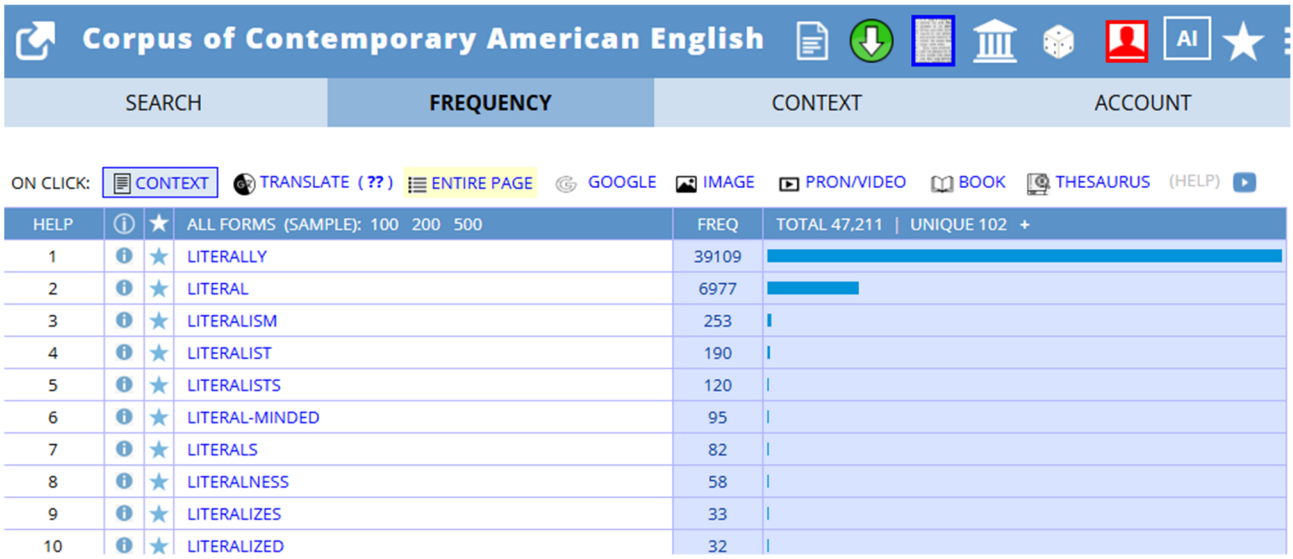
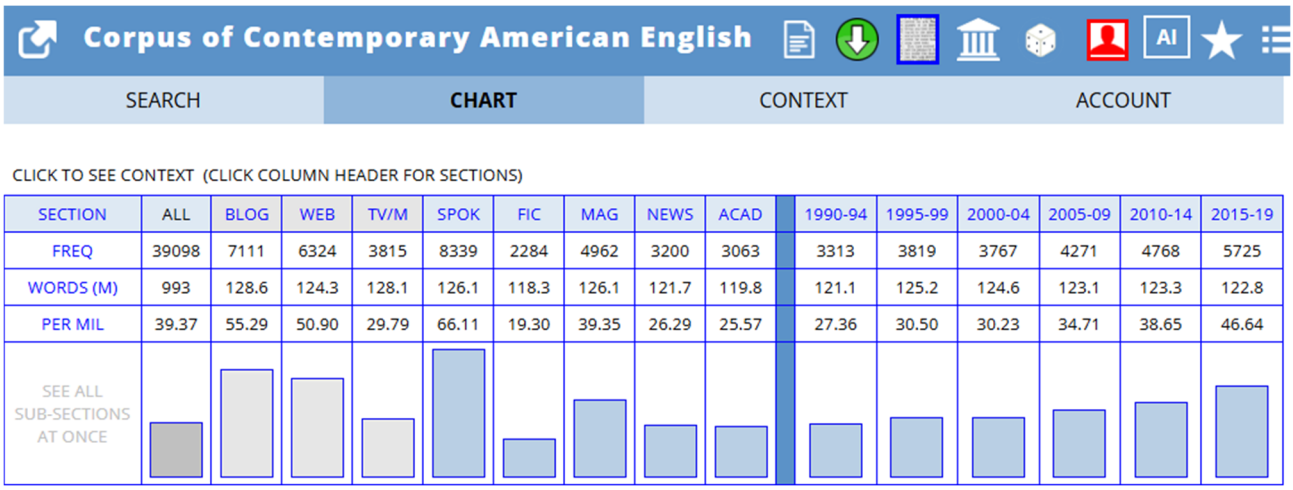


그림 1. List에서 'literal*'를 검색한 결과

2.2. Chart 기능

Chart 기능은 특정 단어나 구문의 시기별(1990-2019) 사용 빈도 변화를 막대그래프 형태로 시각화하여 보여준다. 이는 특정 표현의 사용 빈도가 시간이 지남에 따라 증가했는지 혹은 감소했는지를 한눈에 파악하게 해 준다. 교사나 학생은 이를 통해 언어의 통시적 변화를 추적하거나, 특정 사회·문화적 사건이 언어 사용에 미친 영향을 분석하는 연구를 수행할 수 있다. 왼쪽 하단에 SEE ALL SUB-SECTIONS AT ONCE 를 클릭하면 보다 상세한 결과를 제공해 준다.



SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	39098	7111	6324	3815	8339	2284	4962	3200	3063	3313	3819	3767	4271	4768	5725
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	121.1	125.2	124.6	123.1	123.3	122.8
PER MIL	39.37	55.29	50.90	29.79	66.11	19.30	39.35	26.29	25.57	27.36	30.50	30.23	34.71	38.65	46.64

그림 2. Chart에서 'literally'를 검색한 결과

2.3. Word 기능

Word 기능은 특정 단어 하나를 중심으로 7가지 핵심 정보를 입체적으로 제시한다. 관련 주제어(TOPIC), 연어(COLLOCATES), 관련 단어(RELATED WORDS), 동의어(SYNONYMS), 클러스터(CLUSTERS), 가장 많이 사용된 웹사이트(TEXTS/VIRTUAL CORPORA), 콘코던스 라인(CONCORDANCE LINES)을 제공한다. 특정 단어를 다각적인 측면에서 학습할 때 유용하다. 오른쪽 상단에는 연어(Collocates), 클러스터(Clusters) 등의 추가 기능을 제공한다. Word 기능은 Collocates, KWIC 기능은 연결되어 있음을 확인할 수 있다.

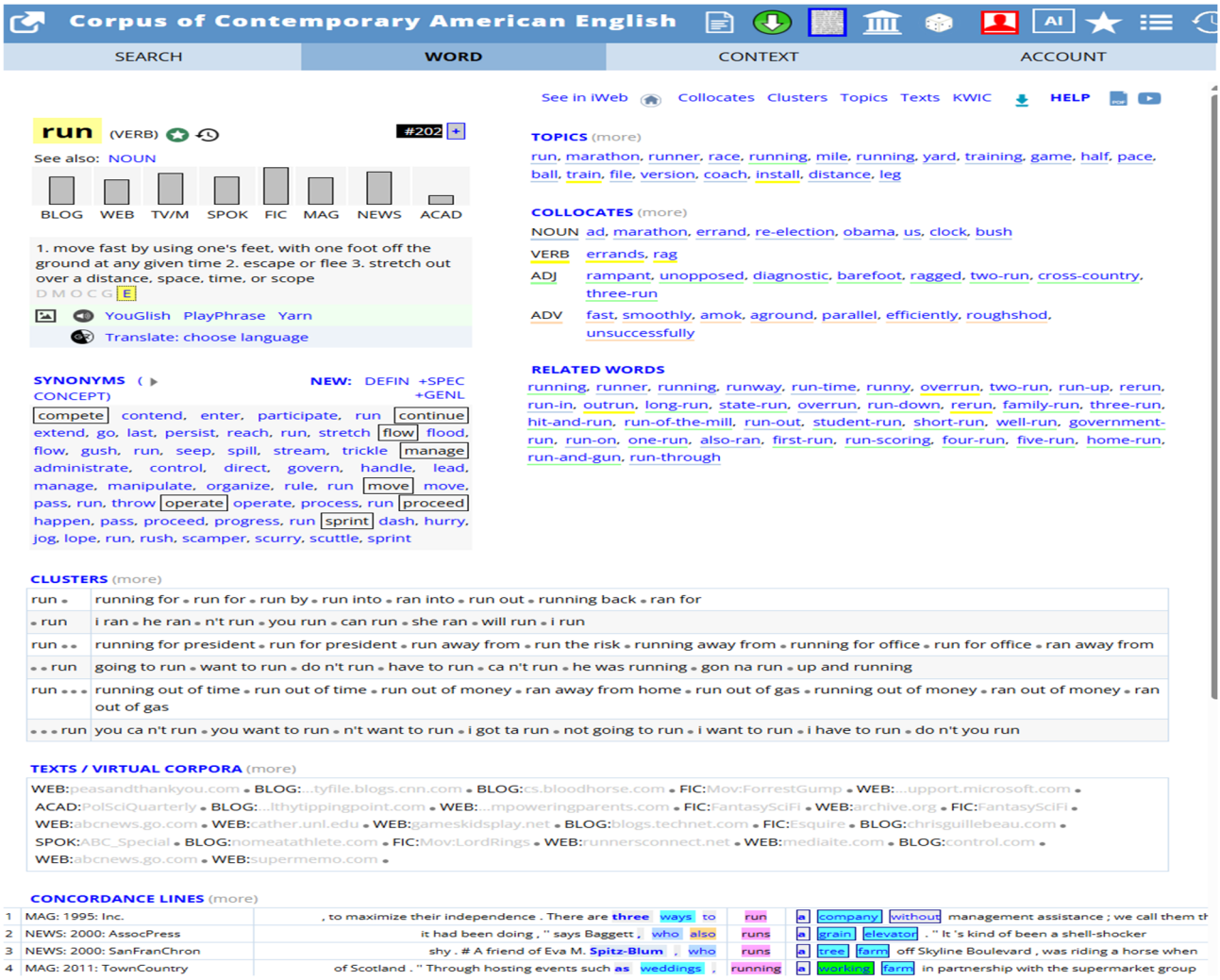


그림 3. Word에서 'run'을 검색한 결과

2.4. Collocates 기능

Collocates 기능은 검색어와 통계적으로 유의미하게 함께 등장하는 단어, 즉 연어(collocates)를 찾아 그 연관성의 강도를 MI 점수(Mutual Information score) 등의 통계치로 제시한다. 이를 통해 연구자는 특정 단어의 의미운용(semantic prosody)이나 관용적인 어휘 결합 관계를 객관적인 데이터에 기반하여 분석할 수 있다. 이 기능은 어휘 연어 학습이나 특정 어휘의 전형적인 사용 환경을 학습시키는 데 매우 유용하다. KWIC를 찾을 수 있어서 맥락 안에서 단어를 학습하는 데 도움을 제공한다.

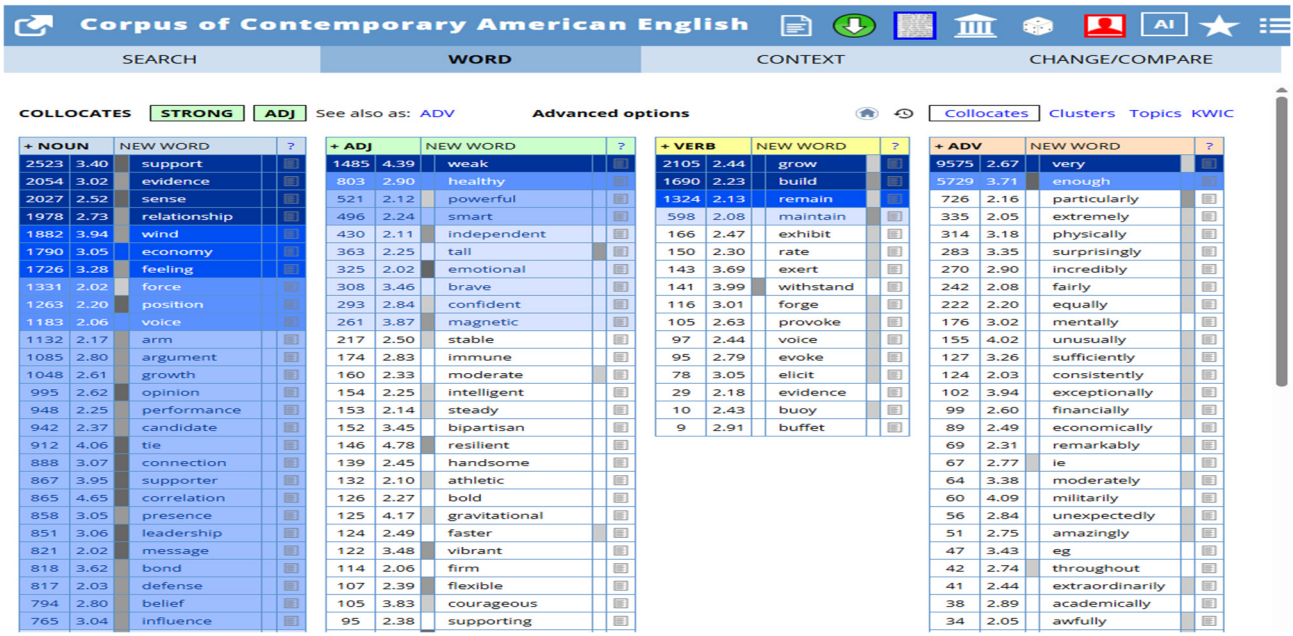
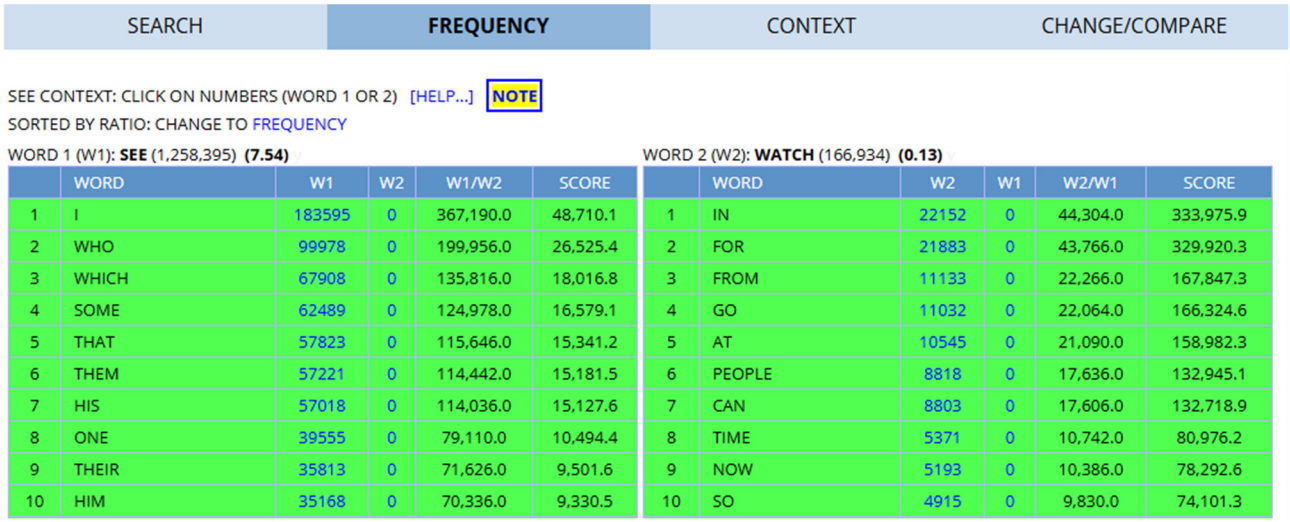


그림 4. Collocates에서 'strong'을 검색한 결과

2.5. Compare 기능

Compare 기능은 두 단어의 연어 관계(Collocates)를 비교하여 의미와 용법의 차이를 파악하는 데 사용한다. 단순히 동의어를 나열하는 사전적 정의를 넘어, 특정 범위에서 단어들이 실제로 어떻게 사용되는지, 그리고 특정 단어와 함께 자주 나타나는 다른 단어들을 통해 해당 단어의 뉘앙스를 심층적으로 분석할 수 있어 어휘 학습에 도움이 된다.



	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	I	183595	0	367,190.0	48,710.1	1	IN	22152	0	44,304.0	333,975.9
2	WHO	99978	0	199,956.0	26,525.4	2	FOR	21883	0	43,766.0	329,920.3
3	WHICH	67908	0	135,816.0	18,016.8	3	FROM	11133	0	22,266.0	167,847.3
4	SOME	62489	0	124,978.0	16,579.1	4	GO	11032	0	22,064.0	166,324.6
5	THAT	57823	0	115,646.0	15,341.2	5	AT	10545	0	21,090.0	158,982.3
6	THEM	57221	0	114,442.0	15,181.5	6	PEOPLE	8818	0	17,636.0	132,945.1
7	HIS	57018	0	114,036.0	15,127.6	7	CAN	8803	0	17,606.0	132,718.9
8	ONE	39555	0	79,110.0	10,494.4	8	TIME	5371	0	10,742.0	80,976.2
9	THEIR	35813	0	71,626.0	9,501.6	9	NOW	5193	0	10,386.0	78,292.6
10	HIM	35168	0	70,336.0	9,330.5	10	SO	4915	0	9,830.0	74,101.3

그림 5. Compare에서 'see'와 'watch'를 검색한 결과

2.6. KWIC 기능

KWIC 기능은 특정 단어가 COCA를 구성하는 여러 하위 장르(예: SPOKEN, FICTION, ACADEMIC 등)에서 얼마나 다른 빈도로 사용되는지를 비교 분석해준다. 교사나 학생은 이 기능을 통해 특정 표현이 구어에서 더 자주 사용되는지, 혹은 학술 텍스트에서 더 자주 사용되는지 등을 정량적으로 확인할 수 있다. 이는 장르별 대화 표현을 학습시키거나 2022 개정 교육과정의 교과목인 미디어 영어와 같이 특정 목적의 영어(English for Specific Purposes, ESP) 교육 과정 설계에서 학습 자료를 활용할 수 있다. 전체적인 맥락을 확인하기 위해 장르를 클릭하면 세부적인 내용을 파악할 수 있어 읽기 교육에도 활용할 수 있다.

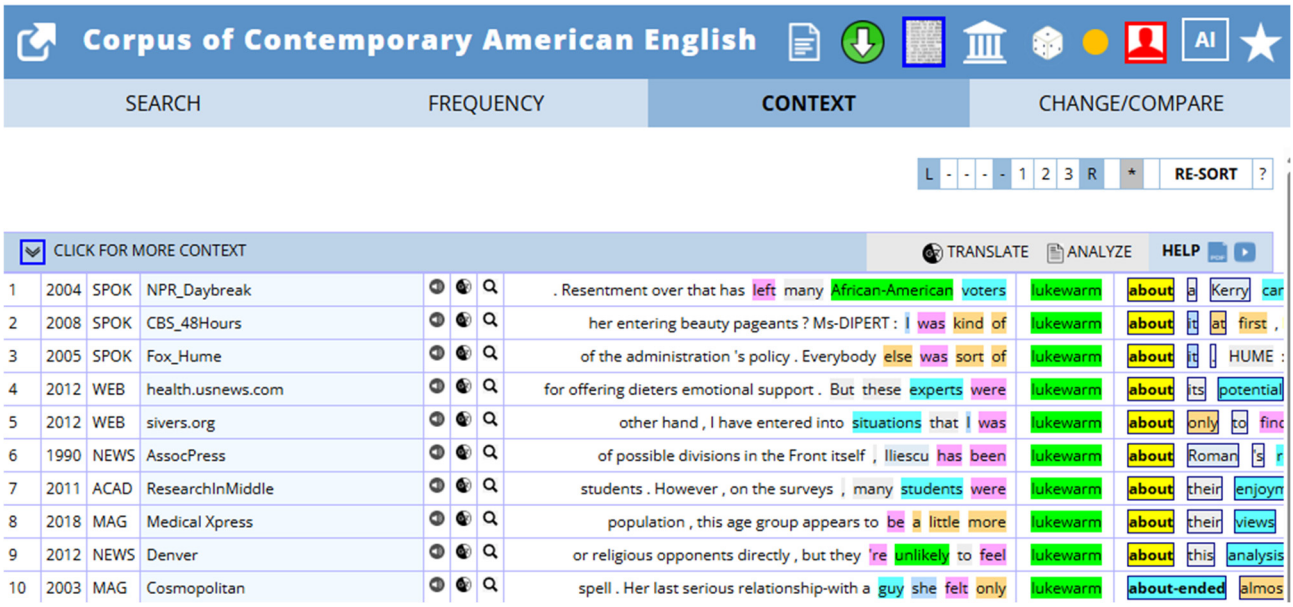


그림 6. KWIC에서 'lukewarm'을 검색한 결과

3. COCA 의 주요 특징

COCA는 단순히 텍스트를 모아 둔 것이 아닌, 체계적인 설계를 통해 다양한 연구 목적에 부합하도록 구축되었다. 구어, 소설, 대중 잡지, 신문, 학술 등 5개의 핵심 장르와 블로그, 웹사이트 데이터를 균형 있게 포함하여, 특정 장르에 편중되지 않은 현대 미국 영어의 전반적인 모습을 보여준다. 이를 통해 교사나 학습자는 장르 간 언어 사용의 차이를 신뢰도 높게 비교 분석하며 교수 학습할 수 있다.

1990년부터 2019년까지 매년 동일한 규모의 데이터를 구축하여, 언어 변화의 추이를 연구할 수 있는 기반을 제공한다. 이는 신조어의 등장과 확산, 특정 문법 구조의 쇠퇴 등을 이해하는 데 활용될 수 있으며 언어에 관심 있는 학생들에게는 흥미로운 탐구 주제가 될 수 있다.

단순히 단어를 검색하는 것 외에도 와일드카드(*), 품사 태그를 활용한 검색, 표제어(lemma) 검색을 이용한 다양한 검색이 가능하고 교사나 학습자가 원하는 언어 데이터를 정밀하게 추출할 수 있도록 지원하여 다각적인 언어 학습에 도움을 준다.

4. 교수 학습 활용 사례 및 한계점

4.1. 교수 학습 활용 어휘, 문법, 쓰기 교육

코퍼스를 활용하면 학술 어휘 뿐만 아니라 어휘 문법적(lexico-grammatical) 의미와 기능까지도 학습시킬 수 있으며 이를 통해 대화, 읽기, 쓰기 기술 향상에도 도움을 준다(김태국, 정채관, 2016; Ivanova, Skitina, Shabanova, & Poliakova, 2023). 특히 Karpenko-Seccombe(2020)는 Compleat Lexical Tutor, BNC—English 코퍼스, Sketch Engine(SkELL), Michigan Corpus of Upper-Level Student Paper(MICUSP)를 사용하여 학생들의 데이터 기반 학습(Data-Driven Learning)을 유도하고 학습자가 스스로 코퍼스 활용법을 학습하고 어휘와 문법 학습과 함께 학술 글쓰기 연계하는 방법을 제시하였다.

이지연(2024)은 영어 예비교사의 영어 쓰기 능력 및 인식에 미치는 영향을 분석하면서 코퍼스를 활용한 영어 교육이 유의미한 결과를 보였다고 하였다. 특히 학술적 글쓰기에서는 Word 기능의 장르 정보, 연어 정보, 클러스터 정보를 활용하면 학습자가 보다 자연스럽고 정확한 영어 표현을 사용할 수 있다. 또한 GSL(General Service List)과 AWL(Academic Word List)의 학술 어휘를 구분하는 기능을 제공하니 학술 어휘 사용에 도움을 제공한다. 또한 Compare 기능을 통해 유사한 의미의 단어들 사이에서 문맥에 적합한 어휘를 선택하는 능력을 기를 수 있다는 장점도 있다.

4.2. 한계점 및 활용 시 유의사항

COCA는 이름 그대로 현대 '미국' 영어를 중심으로 구축되었으므로, 영국 영어, 호주 영어 등 다른 영어권의 언어 사용을 분석하는 데는 한계가 있다. 학생들의 다양한 언어에 대한 이해를 위해서는 BNCweb 과 같은 영국 영어 온라인 코퍼스를 함께 사용하면서 균형을 맞춰 지도할 필요가 있다. 또한 SkELL, MICUSP등과 같은 특정 목적의 온라인 코퍼스도 존재하므로 수업 목적에 맞게 사용하면 된다.

둘째, COCA는 무료 사용자에게 일부 기능과 검색 횟수에 제한을 두며, 모든 기능을 무제한으로 사용하기 위해서는 유료 라이선스가 필요하다. 중·고등학교에서는 해당 비용이 부담될 수 있어 접근성이 제한될 수 있으므로, 모둠 별로 검색할 수 있도록 하는 수업 모델 구안이 필요하다.

마지막으로 COCA와 같은 코퍼스 데이터는 정량적인 통계 데이터와 용례를 제공할 뿐, 그 데이터가 가지는 의미를 해석하는 것은 전적으로 교사나 학습자의 역량에 달려있다. 교사는 먼저 자신의 언어학적 지식과 비판적 사고를 바탕으로 언어 데이터 학습을 시도하고 교실에 적용하지 않는다면 학습자들은 더욱 어려움에 처하게 될 것이기에 체계적인 수업 모델 구안이 그 무엇보다 중요하다고 하겠다.

5. 결론

본 연구는 2022 개정 영어과 교육과정에서 강조하는 인공지능 및 디지털 도구 활용 방향에 부합하여 COCA의 교육적 활용 방안을 탐색하였다. COCA는 10억 어절 규모의 방대한 데이터와 균형 잡힌 장르 구성을 바탕으로 List, Chart, Word, Collocates, Compare, KWIC의 6가지 핵심 기능을 제공한다. 이러한 기능들은 실제 언어 사용 맥락에서 어휘와 문법을 학습할 수 있게 하며, 특히 연어 관계 분석과 장르별 언어 사용 패턴 비교를 통해 학습자의 영어 의사소통 역량 향상에 기여할 수 있다.

COCA와 같은 온라인 코퍼스는 디지털 언어 자원의 교육적 활용은 단순한 기술 도입을 넘어, 학습자가 실제 언어 사용의 맥락 속에서 영어를 탐구하고 학습할 수 있는 새로운 패러다임을 제시한다. 이는 2022 개정 교육과정이 추구하는 자기주도적 학습 능력과 비판적 사고력 함양에도 기여할 수 있어 학교 현장에서 사용할 만한 학습 보조 자료라 하겠다.

참고문헌

교육부. (2022). *영어과 교육과정*. 제2022-33호 [별책 14]. Ministry of Education. (2022). *National Curriculum of English*.

권혁승, 정채관. (2012). *코퍼스 언어학 입문*. 서울: 한국문화사. Kwon, H. S., & Jung, C. K. (2012). *Corpus Linguistics Introduction*. Seoul: Hankook Publishing House.

김태국, 정채관. (2016). 교육적 코퍼스 활용에 관한 학습자의 태도 분석: 고등학생을 중심으로. *중등영어교육*, *9*(1), 47-80. Kim, T. G., & Jung, C. K. (2016). Korean high school student attitudes toward pedagogic corpus use. *Secondary English Education*, *9*(1), 47-80.

이지연. (2024). 코퍼스를 활용한 영어 쓰기 학습 활동이 예비 영어교사의 영어 쓰기 능력 및 인식에 미치는 영향. *교사교육연구*, *63*(4), 489-511. Yi, J. Y. (2024). A study of the effects of corpus-based English writing tasks on prospective English teachers' writing ability and perception. *Teacher Education Research*, *63*(4), 489-511.

Karpenko-Seccombe, T. (2020). *Academic Writing with Corpora: A Resource Book for Data-Driven Learning*. London and New York: Routledge.

Ivanova, E. G., Skitina, N. A., Shabanova, V. P., & Poliakova, N. V. (2023). Corpus-based teaching of specific and academic vocabulary for ESL students. *Journal of Higher Education Theory and Practice*, *23*(19), 151-157.

THE AUTHOR

Taeguk Kim is a teacher at Inje High School in Incheon. He studies corpus linguistics and research on using corpora for educational purposes. He also focuses on the use of artificial intelligence in English language education.

THE AUTHOR'S ADDRESS

First Author
Taeguk Kim
Teacher
Inje High School
12, Gaetbeol-ro, Yeonsu-gu, Incheon, SOUTH KOREA
E-mail: consumption@hanmail.net

Received: 25 July 2025
Received in Revised Form: 5 August 2025
Accepted: 17 August 2025