



빅데이터를 활용한 군사영어 기초어휘목록 개발*

권은영·우경민·진성한·정태영 (육군사관학교)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: December 21, 2022

Revised: January 22, 2023

Accepted: January 30, 2023

Eun-Young Kwon (1st author)
Associate Professor, Dept. of English, Korea Military Academy
eykwon@mnd.go.kr

Kyeongmin Woo (co-author)
Assistant Professor, Dept. of English, Korea Military Academy
martincan@kma.ac.kr

Seonghan Jin (co-author)
Assistant Professor, Dept. of English, Korea Military Academy
shjin0802@gmail.com

Tae-Young Jeong (corresponding author)
Professor, Dept. of English, Korea Military Academy
kma6446@kma.ac.kr

* This paper was supported by 2022 project fund of Future Strategy and Technology Research Center and 2023 research fund of Hwarang-dae Research Institute.

ABSTRACT

Kwon, Eun-Young, Kyeongmin Woo, Seonghan Jin and Tae-Young Jeong. 2023. Development of a basic word list for Military English using big data. *Korean Journal of English Language and Linguistics* 23, 109-130.

The purpose of this study is to show the development process of the Military English Army Basic Word List (MEABWL), which is specifically designed for military English. The process of developing MEABWL involves the following three steps: 1) create a military English corpus (MEA_Corpus) by crawling big data composed of various military-related written and spoken texts; 2) make word lists in three ways (i.e., by using the scope of range, utilizing the corpus comparison approach, and extracting keywords) and select the most appropriate one for the study; and 3) develop a word list incorporating expert judgment in regards to 'familiarity' and 'specialty' based on the word list categorized by the scope of range. By developing and presenting a military English word list through the establishment of a 9,877,104-word military corpus, this study is expected to help learners understand military-relevant texts and educators customize word lists following specific needs in their English for specific purposes (ESP) courses/classes. Finally, the study concludes by offering some suggestions for further research.

KEYWORDS

ESP, military English, word list, big data, corpus

1. 서론

어휘는 영어학습에 있어 중심적인 역할을 한다고 해도 지나치지 않을 것이다. 이에 대해 Wilkins(1972)는 문법 지식이 없더라도 어느 정도 의도하는 바가 전달될 수 있는 것과 달리, 어휘 지식 없이는 원하는 메시지를 전달할 수 없다며 어휘 지식의 중요성에 대해 언급한 바 있다. 이처럼 원활한 영어 구사 능력을 가능하게 하는 데에 있어 어휘 지식이 가지는 역할은 매우 크다고 할 수 있으며, 이러한 어휘 지식의 중요성 및 어휘 교육 효과를 다룬 연구는 다양한 학습자를 대상으로 특수 목적 영어교육 분야에서도 다수 진행되어 오고 있다(강남희 2017, 권선희 2013, 김미영 2019).

영어를 내용적인 면으로 분류하자면 크게 일반 목적 영어(English for General Purposes, EGP) 및 특수 목적 영어(English for Specific Purposes, ESP)로 나눌 수 있다. ESP는 의료, 법률, 비즈니스, 군사 등 전문분야에서 사용되는 영어로서, 이는 또다시 학문 목적 영어인 English for Academic Purposes(EAP)와 직업 목적 영어인 English for Occupational Purposes(EOP)로 나뉜다. 이러한 ESP는 학습자가 특수한 목적을 달성하기 위하여 특정 기술과 영어를 교육하는 것으로서, 그 실용적인 특성으로 인해 EFL 및 ESL 환경에서 활발히 사용되고 있다(권은영 2020, 황필아 2012, Jhang and Parent 2011, Johns and Dudley-Evans 1991).

원활한 학술 목적 영어(EAP) 교육과 관련해 많은 학자들은 언어와 관련된 다양한 기초 지식을 학생들이 갖추고 있어야 한다고 보는데, 특히 외국어로 쓰인 학술적인 글을 읽는 데에 있어 어휘 지식은 그 어떤 언어 지식보다 큰 영향을 미치며(Coxhead 2013, Nation 2001) EAP 학습자의 효과적인 어휘학습을 위한 어휘목록으로 Coxhead(2000)의 Academic Word List(AWL)가 제시되었다. 그러나 비록 AWL이 대학 수학을 위하여 필요한 어휘목록을 제시하는 ESP적 성격을 가지고 있다 해도, 법, 의학 등 좀 더 세분화된 특정 분야에 한정된 전문 어휘를 아우르지 못한다는 한계점이 있어 그간 항공, 화학, 공학 등 전문 영역별 어휘목록을 제공하고자 하는 연구들 또한 진행되어 왔다(이성은 2022, 최인영 2020, Lei and Liu 2016, Valipouri and Nassaji 2013, Ward 1999). 이처럼 특정 전문 영역을 다루는 ESP 학습자들을 위해서는 각각의 특정 분야에 관련될 뿐 아니라 빈번히 사용되는 전문 어휘학습이 병행되어야 하며(정미애 2007), 이는 군사영어 학습에도 적용된다.

현재 우리나라는 그 어느 때보다 복잡하고 어려운 안보환경에 처해 있다. 전작권 전환에 대비한 군 간부들의 영어 소통능력 함양이 미군과의 작전 수행에 필수적인 요소가 될 뿐 아니라, 한미연합 작전과 다국적군 작전을 수행해야 하는 상황 등으로 인해 우리 군에게 있어 ESP적인 군사영어 교수학습의 중요성이 지속적으로 증가하고 있는 것이다(Orna-Montesinos 2013).

실제 야전 교범 및 군 관련 문서를 읽는 데 있어 겪는 어려움 중 하나는 어휘와 관련된 것이다. 이는 군사영어가 (1) 신속 정확한 군사작전에 사용되어야 하는 특수성으로 인해 객관적 사실을 바탕으로 한 간결한 문장이 많이 쓰며, (2) 약어(abbreviations) 및 두문자어(acronyms)를 다수 포함하는 특징을 가지는 것에 기인한다(송근호 2022). 좀 더 자세히 살펴보자면 간결한 문장 사용의 경우, 군사훈련 및 작전 시 무선통신(Radio Communication)을 통한 아군과의 의사소통 과정에서 무전 내용이 잘 들리는지 상대방에게 확인하고자 할 때 “How copy?” 또는 “How do you read me?”이라고 묻는다면, 상대방의 메시지를 잘 이해했다는 표현을 “I have received your

transmission well.” 대신 “Roger,” “Copy,” “Received” 등의 간단한 무선통신 약어로 대체하는 예 등에서 쉽게 찾아볼 수 있다. 또한, 약어 및 두문자어 사용의 경우, 군사작전 및 훈련 이후 해당 작전/훈련이 계획대로 잘 이루어졌는지 등을 검토할 때의 ‘사후검토’를 의미하는 “After Action Report”는 AAR로, 또 공격/방어작전 시 예하 부대에 하달하는 작전명령을 의미하는 “Operation Order”는 “OPORD”로 줄여 사용하는 것뿐 아니라 “CP”와 같은 용어는 군사작전 시 현저한 지형지물(교량, 고층건물, 산 정상 등)에 부여하여 적 또는 아군의 위치를 용이하게 파악할 수 있게 하는 ‘확인점’이라는 “Check Point”를 뜻하기도 하고, 군사작전이나 훈련 간 실제 지휘관이 위치하여 예하부대를 지휘하는 ‘지휘소’라는 “Command Post”라는 뜻도 가지므로 이 같은 두문자어 및 약어들은 해당 상황이나 문맥에 맞게 의미를 파악해야 하는 신중함을 요한다.

이러한 군사영어의 특징을 고려해 볼 때 군사영어 전문 어휘로 구성된 목록을 개발하는 것은 군사 영어교육에 있어 필수 불가결한 작업이라 할 수 있다. 특히 한미연합훈련 및 작전의 교류가 잦고 그에 따른 미군과의 협력 작전 및 임무 수행의 빈도가 높은 한국군의 특성상 군 내부에서는 우선 미군과의 교류에 방점을 둔 어학 능력 향상에 주안점을 두고 있다. 그간 군사영어 소통능력 향상을 목적으로 몇몇 군사영어 관련 코퍼스 연구가 진행되었으나(이한술, 신경식, 이장호 2012a, 이한술, 이장호, 신경식 2012b, Kim 2015, Noguera-Díaz and Pérez-Paredes 2019) 미 육군 대상 군사영어에 한정된 전문 어휘목록은 아직 개발되지 못한 실정이다. 이에 본 연구에서는 미 육군 관련 다큐멘터리, 잡지, 야전 교범 등 다양한 자료에서 추출한 실제적인(authentic) 구어 및 문어를 포함한 빅데이터 자료를 바탕으로 구축한 전문 코퍼스에서 군사영어 ESP 교수학습에 기초적으로 필요하다고 판단되는 어휘를 선별해 목록을 개발하고자 하였다. 즉, 군사영어 분야의 전문 코퍼스 구축을 통하여 군사영어 어휘리스트를 개발, 제시함으로써 본 연구는 향후 효율적인 ESP 수업과 함께 학습자의 군사영어 소통능력 함양에 도움을 주는 것을 목적으로 한다.

2. 이론적 배경

2.1 군사영어의 특징 및 중요성

군사영어는 군(military)이라는 특수한 조직 또는 환경에서 주로 사용되는 특수 목적용 영어(ESP)의 한 부류로서, Dudley-Evans와 St. John (1998)이 분류한 전문 영역에 따른 ESP의 분류 유형 중 특수직업을 위한 영어(EOP)에 속한다. 즉, 군사영어는 일반적인 상황이나 문맥에서는 잘 활용되지 않는 특정한 ‘군사적인’ 의미가 있는 명사, 형용사, 동사 및 다양한 종류의 약어, 두문자어, 작전 및 전술 용어 등을 포괄한다. 이러한 군사영어는 일반적인 상황에서 쓰이는 영어와는 달리 군이라는 특수한 환경에서 특정한 목적(전술훈련, 작전계획 수립, 무선통신 등)을 달성하기 위해 반드시 ‘학습’이 되어야만 원활한 의사소통을 가능하게 한다는 특징이 있다. 또한, 신속하고 정확한 의사소통의 목적을 달성하기 위해 간결한 문장을 사용하고(송근호 2022), 특정한 상황에서 그에 맞는 의미를 지니는 어휘(예를 들면, friendly fire)를 사용한다는 특징 또한 가진다.

군사영어의 중요성은 크게 두 가지 측면에서 살펴볼 수 있는데, 첫 번째로는 국제적인 군사적 환경에서 다양한 국가의 군 구성원들이 소통하는 ‘국제언어로서의 영어(English as an

International Language)’라는 점이다. 이미 세계 여러 분쟁지역에서 두 국가 또는 그 이상 국가들의 군이 평화강제작전 및 평화유지작전 등과 같은 다양한 국제적인 또는 다국적 군사 임무를 수행하고 있고(Pierce and Dixon 2006), 그러한 국제적인 군사적 환경에서 영어는 매개어(vehicular language) 또는 공용어(lingua franca)로서 다국적군 구성원들이 원활한 의사소통을 할 수 있는 가장 중요한 언어적, 문화적 수단이다(Orna-Montesinos 2013). 최근 러시아-우크라이나 전쟁으로 그 중요성을 더욱 주목받고 있는 북대서양 조약기구(North Atlantic Treaty Organization; NATO)의 구성원들도 영어를 공용어로 쓰고 있다(김나래, 이한솔, 정태영 2019). 이러한 점에서 볼 때, 군사영어는 다양한 국적과 문화적 배경을 가진 군 구성원들이 국제적인 군사 임무를 성공적으로 달성할 수 있도록 유효한 구실을 하는 “군사 공용어(English as a military lingua franca)”라고 할 수 있다(진성한 2022, p. 61).

두 번째로, 한반도의 지정학적 위치와 연계하여 한미연합훈련의 중요성이 날로 증대되고 있는 가운데, 군사영어는 유사시 한미연합사령부의 작전 통제 하 우리 군이 미군과의 연합작전을 수행할 때 필수적인 요소이다. 또한, 향후 전시 작전통제권 전환에 대비하여 한국군이 주도적으로 한미 연합훈련 및 연합작전 등을 이끄는 데에 있어 군사영어를 구사하는 능력을 배양하는 것은 매우 중요하다. 여기서 군사영어 능력을 향상시킨다는 것은 단순히 미군과의 의사소통 측면에서 회화능력을 증진하는 것뿐만 아니라, 한미동맹이라는 기치 아래 군사영어를 통해 연합작전을 계획 및 지도하고, 미군과의 효율적인 작전수행능력을 기르는 등 한미 연합방위를 실현하기 위한 실질적인 능력을 향상시키는 것을 의미한다(송근호 2022). 또한, 해외파병을 통한 국제평화유지활동, 해외 군사교육, 해외 다국적 연합훈련 등 다양한 국제적 기회에서 한국군이 외국군들과 소통하는데 있어 군사영어는 필수불가결한 요소이다.

이렇듯, 군사영어 교육 및 학습에 대한 중요성에도 불구하고 이와 연관된 연구들은 많지 않은 것이 사실이다. 실제 군사영어는 해방 이후 미 군정 시기 군사영어학교에서 장교후보생들에게 영어교육을 했던 것을 시초로 거슬러 올라갈 만큼 역사가 오래되었으나, 군사영어가 군이라는 특수한 집단이 특정한 상황에서 사용한다는 제약으로 인해 충분한 학문적 관심을 얻지 못하였다. 그러나 국내에서는 군사영어를 가르치는 사관학교 및 국방어학원 등 관계 기관에서 군사영어를 어떻게 효과적으로 학습하고, 교육하고, 평가할 것인가에 대해 꾸준한 관심을 경주해 오기는 하였으나(신경식, 박지환 2009, 정태영, 신경식, 허진 2009, 김나래 외 2019, 이한솔, 김나래 2020, 이한솔, 오준혁, 정태영 2021). 아직 군사영어의 교육 및 학습에 대한 정량적인 학습효과나 효율적인 학습방법 및 평가방법에 관해서는 연구가 많이 부족하므로 향후 꾸준한 연구적 관심이 요구된다.

2.2 코퍼스 관련 선행연구

코퍼스(Corpus)는 언어 연구, 그중에서도 특히 어휘연구에 주로 활용되는 방대한 양의 문어 및 구어 텍스트들의 집합이다(강범모 2003). 코퍼스는 컴퓨터 활용으로 인해 많은 양의 정보를 단시간 내에 분석해 주는 것뿐 아니라 통계적 분석을 통해 특정 기준(빈도수 등)에 맞는 정보를 추출함으로써 학습자들에게 유용한 지표를 제공할 수 있다는 장점을 가지며(Biber, Conrad and Reppen 1998), 이로 인해 그간 코퍼스를 활용한 EFL 학습 관련 연구들이 다수 진행되어 왔다(김성식 2007, 김태국, 정채관 2016, 이경은, 신정화 2017, 이효진, 민덕기 2016, 임수영, 이은주

2012, 정채관 2021, 한나래, 이수화 2009).

이처럼 코퍼스는 EFL 환경에서 언어 교수 및 학습에 유용하게 사용될 수 있다는 장점으로 인해 관련 연구들이 꾸준히 진행됐지만, 군사영어의 학습효과를 높일 수 있는 코퍼스 활용 국내 연구는 많이 부족한 실정이다. 그럼에도 불구하고 군사영어 학습에 관계된 몇몇 코퍼스 연구가 보고된 바 있는데, 일례로 신경식과 박지환(2009)은 군 교범 중심으로 구성된 읽기 및 쓰기 목적의 코퍼스를 구축한 바 있으며, 정태영(2018)은 국내 학습자들의 구어 데이터를 활용하여 학습자 구어 영어 코퍼스를 구성한 후 이를 실제 수업에 활용할 수 있는 방안을 제시하였다. 이한솔 외(2012a, 2012b) 역시 각각 실용자료에 근거해 군사영어 코퍼스를 한미 연합작전에 실제로 활용할 수 있는 방법 등을 연구하였는데, 좀 더 구체적으로 이한솔 외(2012a)는 기존의 교범 중심 코퍼스에 추가하여 말하기 중심의 자료들을 보완함으로써 좀 더 실질적으로 군사영어 회화상황에서 쓰일 수 있는 코퍼스를 구축하고자 하였으며, 이한솔 외(2012b)는 실용적인 말하기 위주의 코퍼스를 구축하여 실제로 군사영어에서 사용되는 어휘와 가장 잘 조합되는 어휘 및 빈도수 등을 파악하였다. 이와 같은 ESP 코퍼스 구축은 해당 분야 어휘리스트 작성을 위한 기초작업이 된다.

2.3 어휘목록 관련 선행연구

영어를 외국어로 배우는 EFL 환경의 학습자들에게 있어 가장 기본적이면서도 필수적인 요소는 어휘 습득이다(Saville-Troike 2006). 그동안 학습자들의 효율적인 학습 보조도구로서 다양한 어휘목록이 개발되어 왔는데(Coxhead 2000, Xue and Nation 1984, West 1953), 이 중 가장 대표적인 어휘목록으로는 West(1953)가 개발한 General Service List(GSL)이 있다. GSL은 총 250만 단어로 구성된 코퍼스를 기반으로 추출한 2,284개의 표제어로 구성된 어휘 목록으로서, 현재 사용되는 어휘를 반영하고 있지 않다는 점에서 비판을 받고 있긴 하지만(Carter 2012), 텍스트 포괄률이 약 80%를 차지하기 때문에 Nation과 Hwang(1995)은 EAP 어휘 학습 이전에 공부하기에 효과적이라고 주장하였다.

GSL 이후 대표적인 어휘목록으로 Coxhead(2000)가 제시한 Academic Word List(AWL)가 있다. AWL은 법(law), 과학(science), 상경(commerce), 예술(arts)의 4가지 학문 영역 및 28개 세부 교과목 텍스트의 총 350만 개 토큰으로 구성된 문어 학술 코퍼스에서 추출한 570개의 단어(word family)으로 이루어진 EAP 어휘목록이다. 여기에서 주목할 점은 AWL은 GSL을 이미 학습했다는 전제하에 개발되었기 때문에 GSL에서 다루는 어휘는 중복하지 않고 있다는 점이다. AWL은 전공에 무관하게 다양한 학문 분야에서 공통적으로 사용되는 어휘를 목록화한 것으로서 학술 텍스트 포괄률은 약 10%이다(Nation 2001).

AWL 개발 이후로도 ESP 분야에서 다양한 어휘리스트가 개발되어 왔다. 일례로 의학 분야에서는 Wang, Liang과 Ge(2008)이 의학연구논문 288편에서 1,093,011개의 의학 분야 학술 어휘를 수집하여 코퍼스를 구성한 후 관련 어휘목록인 Medical Academic Word List를 작성해 분석하였고, 이와 유사하게 Yang(2015)은 간호 관련 연구논문 252편을 활용하여 만든 코퍼스를 기반으로 어휘목록을 만들었으며, 국내에서는 이성은(2022)이 보건의료 계열 전공 대학생들의 글로벌 의사소통역량 강화를 목표로 이를 위하여 그녀는 TED 내 의학 관련 강연 251개의 대본을 수집해 총 512,013개의 토큰으로 구성된 코퍼스 기반의 의학영어 어휘목록을 개발하였다.

의학 이외에 비즈니스, 해양, 항공, 기계공학 등의 분야에서도 관련 연구가 진행되어 왔다. 먼저, 정미애(2007)는 비즈니스 분야 학술지 논문, 경제 서적, 전문지 기사 등 비즈니스 분야 출판 및 보조자료로 구성된 코퍼스에서 얻은 관련 어휘목록을 제시하였고, 장세은과 이성민(2013)은 부산광역시교육청과 한국교육과학기술부에서 각각 출판한 고등학교 해양영어 교과서 두 권을 기반으로 코퍼스를 구축한 후 해양영어 관련 어휘리스트를 개발하였으며. 김미영(2019)은 구어 항공관제영어 코퍼스인 ATCOSIM과 두 개의 참조 코퍼스인 BNC Sampler Spoken 및 COCA Spoken에서 각각 고빈도 어휘 리스트를 추출해 항공관제 영어의 특징을 비교하였다. 이외에도 기계공학 분야의 영어 논문초록으로 코퍼스를 구성하여 전문 어휘를 추출한 신창원(2012) 등 전문 분야 관련 ESP 특수 소규모 코퍼스를 구축하고 관련 용어를 추출하여 제시한 연구가 점차 증가하고 있다.

이와 같은 시도는 군사영어 분야에서도 찾아볼 수 있다. 우선 신경식과 박지환(2009)은 연합작전에 미군과 연합군이 실제 사용하는 교범을 활용하여 군사영어 전문 코퍼스를 구축하여 군사영어 전문용어 및 연어 목록을 제시하였으며, Kim(2015)는 해군 관련 저서, 논문, 매뉴얼, 연설문 등 문어와 구어 텍스트로 구성된 Military English Corpus를 구축한 후 이곳에서 추출한 고빈도 어휘 20개 리스트와 FROWN 코퍼스에서 추출한 고빈도 어휘 20개 리스트의 직접 비교 및 장르별 고빈도 어휘 리스트 비교를 통해 해군 영어 표현의 특징을 분석하였다.

이처럼 의학, 해양, 항공, 군사 등 다양한 영역에서 특수 목적 영어의 성격을 고려한 어휘목록이 제시되고 있음을 알 수 있으나, 상기 연구들은 ESP 코퍼스에 기반하여 고빈도 어휘리스트를 추출하였다는 공통점을 가진다. 즉, 어휘선정 기준이 빈도수(frequency)에 국한된다는 것이다. 이에 반해 신동광과 주현우(2008)는 빈도수 이외에 사용범위(range), 사용분포(frequency dispersion)의 기준을 적용하여 단어 기반의 어휘 검색 프로그램을 개발함으로써 2006년 발표된 영어과 개정 교육과정을 따르는 교과서 내 어휘들을 검색할 수 있게 하였다. 한 걸음 더 나아가 이문복과 신동광(2015)에서는 2008년의 연구와 달리 사용분포 대신 어휘 친숙도(familiarity)를 적용함으로써 한국인 학습자의 학습 특성을 고려한 어휘검색 프로그램을 개발해 제시하였다. 본 연구는 신동광과 주현우(2008) 및 이문복과 신동광(2015)을 참조하여 빈도수, 사용범위, 사용분포와 전문성을 적용하여 군사영어 ESP 어휘리스트를 개발, 제시하고자 한다.

3. 단어 목록 개발절차

3.1 크롤링을 통한 코퍼스 구축

크롤링이란 대량의 텍스트를 비교적 간단한 코드를 이용하여 자동으로 수집하게 해주는 기술이다. 이러한 크롤링은 스크래핑(scraping), 크롤링(crawling), 파싱(parsing)으로 세분되며, 이는 텍스트 자동 수집 기술에 있어 세밀한 정도의 순서에 따른 분류이다(장철원 2022). 본 연구의 경우, 코퍼스 구축 과정에 있어 각 소스 별로 원하는 특정한 부분(예. 광고를 제외한 텍스트, 이미지를 제외한 텍스트 등)만을 의도적으로 수집하는 만큼, 엄밀히 말해 파싱을 통하여 코퍼스를 구축했다고 할 수 있다.

본 연구에서는 파이썬으로 코드 작성 및 실행 이후, 구글의 크롬 셀레니움(Selenium)을 사용하여 모든 파싱 작업을 시행하였으며, 웹사이트별 수집 방법의 상이성을 해결하기 위하여 각기 다른 방식으로 문어 텍스트와 구어 텍스트 수집하였다. 즉, 문어 텍스트 수집의 경우, 각 뉴스 기사 웹사이트에 맞는 코드를 작성하였고, 구어 텍스트 수집을 위해서는 동영상 스트리밍 웹사이트에 맞는 코드를 작성하였는데, 구어 텍스트 수집의 경우, 좀 더 세부적으로, 우선 각각의 기사나 스트리밍 페이지의 주소(URL)를 수집하였다(그림 1 참고).



그림 1. 링크 수집

비디오 스트리밍 사이트에서 수집한 텍스트는 대개 음성인식 기술의 하나인 Speech-to-Text(STT) 기반의 자동자막생성 시스템을 활용하였다. 그러나 오래전에 업로드된 영상 등 자료수집 과정 중 자동자막 기능이 비활성화된 경우가 있어, 그러한 영상들의 자막은 본 연구의 텍스트 수집에서 제외하였다. 다양한 방식으로 수집한 모든 텍스트는 각각 전처리 시행 이후 대규모 코퍼스인 Military English Army(이하 MEA) Corpus를 구축하였으며, 이를 위해 사용된 텍스트의 구성은 아래 표 1과 같다.

표 1. 군사영어 전문 코퍼스(MEA Corpus) 구축에 사용된 텍스트

텍스트	상위 카테고리	하위 카테고리
	미 육군 교범	
문어 텍스트	미 육군 관련 웹사이트	1) military.com 2) army.mil
	미 육군 잡지	1) ARMY Magazine(by Association of US Army)
구어 텍스트	비디오 스트리밍 사이트 내 미 육군 관련 채널	1) 미 육사 채널
		2) ROTC--Univ. of IOWA
		3) USARMY
		4) military.com
	미 육군 관련 다큐멘터리	1) Medal of Honor 2) Surviving the Cut

3.2 군사영어 어휘리스트(MEABWL) 선정 기준

코퍼스를 기반으로 하는 어휘선정 기준에는 통상 빈도수(frequency), 사용범위(range), 분포 빈

도수(dispersion)가 적용되지만(신동광 2011), 본 연구에서는 이문복과 신동광(2015)에서와 같이 분포 빈도수와 빈도수 간의 높은 상관도를 이유로 분포 빈도수를 선정 기준에서 제외하였다. 대신 특수 목적 영어라는 성격을 고려해 전문성(specialty)이라는 기준을 추가 적용하기로 하였으며, 본 연구에서 군사영어 기초 어휘목록(Military English Army Basic Word List, MEABWL)을 만들기 위하여 적용한 어휘선정 기준의 정의 및 의미는 다음과 같다.

3.2.1 사용범위(Range)

하나의 어휘는 다양한 코퍼스에서 사용될 수 있는데, 이를 측정하는 어휘선정 기준이 바로 사용범위이다. Coxhead(1998, 2000)은 대학 전공 코퍼스 28개 중 최소 15개 전공에서 사용되는 어휘를 추출하여 AWL을 개발하였다. 본 연구에서는 기존 ESP 코퍼스 8개를 비교 코퍼스로 사용하였으며, 이는 예술(Arts), 상업(Commerce), 법(Law), 과학(Science) 분야의 문어 코퍼스(written corpus) 4개와 인문학(Arts & Humanities), 사회과학(Social Sciences), 자연과학(Physical Science), 생명 의료과학(Life & Medical Sciences) 분야의 구어 코퍼스(spoken corpus) 4개로 구성되어 있다.

3.2.2 빈도수(Frequency)

선행연구에서 살펴보았듯이 대다수의 코퍼스 기반 어휘목록은 빈도수 순에 따라 어휘가 추출되었으며, 이러한 경향은 다양한 분야의 특수목적 영어 어휘목록 개발에도 다수 적용돼 오고 있다(권선희 2013, 김미영 2019, 이성은 2022). 즉, 빈도수는 어휘목록 제작에 가장 보편적인 기준이라 할 수 있다.

3.2.3 친숙도(Familiarity)

비원어민 학습자는 원어민과 동일한 빈도의 어휘에 노출되기 쉽지 않다. 따라서 친숙도는 어휘 등급을 구분할 때 비원어민 학습자의 특성이 고려된 기준으로서 종종 사용되어 왔다(신동광 2011, 이문복, 신동광 2015). 본 연구에서는 이문복, 신동광(2015, p. 122)에서 개발된 측정 척도를 활용하여 친숙도 정도를 살펴보았다.

표 2. 어휘 친숙도 측정 척도(이문복, 신동광 2015, p. 122)

척도	기준 설명
5	표현의 사용이 매우 친숙함
4	표현의 사용이 비교적 친숙함
3	표현을 잘 사용하지 않으나 의미는 잘 알고 있음
2	표현을 거의 사용한 적 없으나 의미는 대략 알 것 같음
1	표현을 듣거나 본 적 없어 의미를 전혀 알지 못함

3.2.4 전문성(Specialty)

전문 어휘(technical vocabulary)는 어느 특정 분야, 전공, 또는 주제에 관계한 어휘이다(이제영, 김정렬 2013). Nation(2001)은 전문 어휘선정에는 특정 전공 및 분야에서 자주 사용되는 어휘의 형태 및 의미 고려가 중요하며, 연구자의 직관에 따라 기준점이 결정된다고 언급하였다. 본 연구에서는 단어의 형태와 의미를 중심으로 전문 어휘를 산출하고자 Nation(2001, pp. 198-200)의 전문 어휘선정 기준척도를 각색하여 다음 표 3과 같이 전문성 척도를 개발하였다.

표 3. 어휘 전문성 측정 척도(Nation, 2001, pp. 198-200 각색)

척도	기준 설명
3	군사적 ¹ 인 분야 외에는 단어의 형태가 좀처럼 등장하지 않는 경우
2	단어의 형태가 군사적 분야 내외에서 모두 사용되는 경우
1	단어의 형태가 비군사적인 분야에서 더 자주 사용되며, 의미에도 큰 차이 없는 경우

3.3 군사영어 어휘 선정 기준 적용 절차

군사영어 전문 코퍼스(MEA Corpus) 활용 군사영어 단어 어휘 목록(MEABWL) 개발을 위한 어휘 선정에 본 연구에서는 3.2에 제시된 세 가지의 기준으로 세 차례에 걸쳐 각각의 기준을 적용해보았다.

3.3.1 1차 기준 적용 단계: 사용범위

Range_Familizer²⁵는 British National Corpus의 음성 언어 기준, 20,000 단어 내의 모든 단어를 단어군(word family)으로 변형, 제시한다. 그뿐 아니라 이 프로그램은 어휘의 품사에 무관하게 굴절 및 파생 변이형을 포함해 제시하므로 어휘리스트 제작을 용이하게 해주는 장점이 있다(신동광, 주현우 2008).

1차 기준 적용 단계에서 본 연구는 MEA Corpus 내 어휘가 ESP 코퍼스 8개와 비교해 볼 때 겹치는 사용범위(range)가 어느 정도인 지에 대해 Range_Familizer²⁵ 프로그램을 활용하여 알아보았다. 예를 들어, MEABWL 내 단어의 어휘(키워드) 한 개가 8개의 전문 코퍼스 중 어떤 것과도 겹치지 않는 경우, 해당 키워드의 결과값은 숫자 1이 되지만, 겹치는 코퍼스가 한 개 있다면 결과값은 2가 된다. 반면, 전문 코퍼스 8개 모두에 출현하는 키워드가 MEA Corpus에도 있는 경우, 해당 키워드의 결과값은 9이다. 본 연구에서는 사용기준(range)을 선정 기준으로 적용함에 따라, 결과값이 1인 키워드는 군사영어에서만 쓰일 확률이 높은 것으로, 반면 결과값이 9인 키워드는 그 반대인 비군사적, 즉 보편적인 영어단어로 보았다.

¹ 본 연구에서 '군사적'이라는 용어는 병영 생활, 군내 활동, 교리, 장비, 학문(군사사, 국제관계, 무기시스템 공학 등) 등에서 사용되는 것을 의미한다.

3.3.2 2차 기준 적용 단계: 빈도수

본 연구의 2차 기준 적용 단계에서는 빈도수(frequency)도 함께 포함되어 결과가 제공되었다 (그림 2 참고). 빈도수는 코퍼스 내 특정 단어가 얼마나 쓰였는지 알려주는 지표이므로 MEA Corpus 내에 나타난 해당 키워드의 중요도를 알아볼 수 있다.

Types Found In Base List One											
TYPE	RANGE	FREQ	F1	F2	F3	F4	F5	F6	F7	F8	F9
A	9	194336	82189	7593	8757	9717	19072	9677	23113	15694	18609
ABACK	3	4	2	0	1	1	0	0	0	0	0
ABACUS	1	1	0	0	0	0	0	0	0	0	1
ABALONE	1	1	1	0	0	0	0	0	0	0	0
ABANDON	8	69	27	3	7	6	3	0	11	3	9
ABANDONED	8	181	59	0	17	3	22	5	35	23	17
ABANDONING	6	34	7	0	2	0	1	0	12	4	8
ABANDONMENT	7	63	5	2	0	6	12	0	18	8	12
ABANDONS	5	13	2	0	4	2	0	0	4	0	1
ABASED	1	1	0	0	1	0	0	0	0	0	0
ABATE	3	3	1	0	0	0	0	0	1	1	0
ABATED	2	5	1	0	0	0	0	0	0	4	0
ABATEMENT	5	29	10	0	0	10	1	0	3	5	0
ABATES	1	2	0	0	0	0	0	0	0	2	0

그림 2. 리스트에 있는 단어들

3.3.3 3차 기준 적용 단계: 친숙도와 전문성

3차 기준 적용 단계에서는 군사영어를 가르친 경력이 있는 세 명의 평가자(rater)에게 전문성 (expert judgment)을 기반으로 1부터 5까지의 리커트(Likert)척도를 활용해 군사영어 기준으로 어휘의 친숙도를 평가하도록 하였다(그림 3 참고). 평가자들은 모두 군사영어 과목을 가르친 경험이 있지만, 교수 경력 및 군사훈련 경험의 상이성으로 인해 군사영어 어휘 지식의 정도에 있어 다소 차이가 있었다. 따라서 본 연구에서는 세 명의 점수 차가 너무 큰 키워드는 협의를 통해 조정하는 절차를 거쳤으며, 최종적으로 Cronbach Alpha 값을 이용해 평가자 세 명이 도출한 결과값에 대한 평가자 간의 신뢰도(interrater-reliability)를 조사하였다.

	BASE ONE FAMILIES	RANGE	Frequency	Rev 1	Rev 2	Rev 3
1	INFANTRY	1	2808			
2	TRUMP	1	944			
3	HOWITZER	1	792			
4	SNIPE	1	624			
5	ORDNANCE	1	576			
6	TWITTER	1	545			
7	CHAPLAIN	1	474			
8	CONVOY	1	378			

그림 3. 평가자에게 제공된 엑셀 파일

친숙도(familiarity) 정도와 별도로 군사영어 전문가 3명을 대상으로 전문성(speciality) 정도 또한 평가하였는데, 전문성 정도는 1부터 3까지의 척도를 사용했으며, 여기에서 1은 가장 일반적인 단어를, 2는 군사영어와 일반적인 영어 사용범위에 모두 들어가는 카테고리의 단어를, 그리고 3은 군사영어에서 쓰이는 뜻이 지배적인 단어를 의미한다.

본 연구는 친숙도(familiarity)와 전문성(specialty)에 해당하는 점수를 획득한 이후 전문성 정도를 세 개의 항목(category)으로 분류하였으며, 이를 친숙도 값 기준으로 정렬하였다. 그 결과, 친숙도 평균점수가 가장 높은 5를 받은 단어 어휘들 모두 하나도 빠짐없이 항목 3에 해당하였고,

항목 2에는 차순위 친숙도 정도를 나타내는 값들이 순서대로 나타났으며, 항목 1에는 가장 군사 어휘와 거리가 멀거나 기능어 중심의 단어 어휘들이 분류되는 것을 확인할 수 있었다.

아래 그림 4에서 볼 수 있듯이, 본 연구에서는 군사영어 기초 어휘목록 개발을 위해 Range 프로그램을 활용하여 항목 1과 항목 2에 해당하는 단어들을 선택하여 BASEWRD1과 BASEWRD2의 디지털 파일로 만들고, 텍스트 포괄범위(text coverage) 또한 알아보았다.

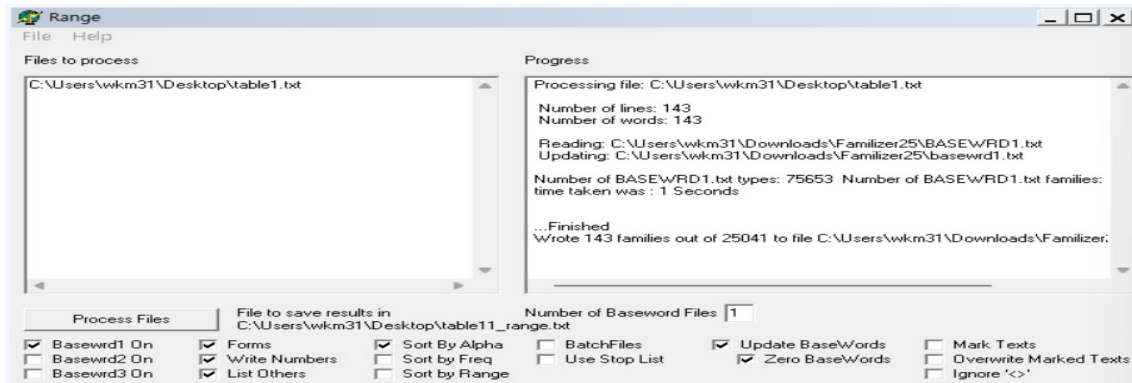


그림 4. Range 프로그램을 통한 어휘목록 개발

3.4 소프트웨어 프로그램 활용 MEABWL 생성 방법

본 연구에서 군사영어어휘목록을 생성하기 위하여 사용한 소프트웨어는 Range_Familizer25와 AntConc 프로그램이다. 여기에서 상기 소프트웨어 프로그램을 활용하여 MEABWL이 생성하는 데에 사용한 방법은 총 세 가지로서, 적용 범위(range) 활용법, 코퍼스 비교법, 그리고 키워드 추출법이 이에 해당한다.

3.4.1 Range 분류법

표 4는 MEA Corpus와 8개의 ESP 전문 코퍼스를 대상으로 각 단어 키워드를 Range(사용범위) 1부터 9까지 중 Range_Familizer25를 활용하여 해당하는 번호를 나타낸 결과를 보여준다. 빈도수를 기준으로 상위권에 나타난 키워드 중, Range 값이 작을수록(1에 가까울수록) 군사적인 맥락의 텍스트에서 흔히 보이는 단어 목록이 추려진 것을 알 수 있다. 또한, Range 9의 최상위권 빈도의 키워드들은 9개의 코퍼스 모두가 가지고 있는 단어 목록이 추려졌기 때문에, 영어에서 가장 흔히 볼 수 있는 기능어들이 관사와 전치사가 다수 나타난 것을 확인할 수 있다. 그러나 Range 값이 작은 키워드들 모두 군사어휘라고 보기는 어렵다. 이는 연구자들이 구축한 MEA Corpus에는 군사/군대와 관련된 여러 텍스트 자료가 섞여 있기 때문인데, 이로 인해 Trump(2위), Twitter(6위)와 같은 단어들이 Range 값 1을 가질 뿐 아니라 상위빈도에 포함되어 있는 것을 확인할 수 있다. 다시 말하자면, 상위빈도 키워드들 중 Range의 값이 낮으면 낮을수록 MEA Corpus에서만 나타나는 군사적인 어휘들이 대거 포함되어 있으며, Range 값이 9에 가까울수록 군사적인 것과 거리가 먼 어휘들이 포함돼 있으나, 이는 순전히 빈도수와 Range 값을 이용하여

일차적으로 분류한 결과이므로, 다른 방법들을 추가적으로 시도해 보아야 할 것으로 보인다.

표 4. Range 범위별 최다 빈출 단어 리스트

사용범위 순위	1	빈도	2	빈도	3	빈도
1	INFANTRY	2808	BATTALION	5836	DECONTAMINATE	2090
2	TRUMP	944	PLATOON	3197	MUNITION	1491
3	HOWITZER	792	ARTILLERY	2646	MEDAL	1185
4	SNIPE	624	RECONNAISSANCE	2446	RANGER	886
5	ORDNANCE	576	MATERIEL	1173	RADIOLOGY	874
6	TWITTER	545	SOP	664	REGIMENT	857
7	CHAPLAIN	474	AZIMUTH	533	MARINES	533
8	CONVOY	378	SQUADRON	351	WARRIOR	500
9	CADET	324	INSURGENT	345	MEDIC	442
10	TALIBAN	316	DISMOUNT	320	TOW	404
사용범위 순위	4	빈도	5	빈도	6	빈도
1	SERGEANT	5000	BRIGADE	5948	COM	4670
2	AMMUNITION	2237	TROOP	3057	CORPS	4522
3	MARINE	1737	FORT	2938	CIVILIAN	3672
4	SYNCHRONIZE	1396	VETERAN	1350	LOGISTICS	2632
5	ARMOUR	1226	MAJORS	911	TERRAIN	2326
6	HELICOPTER	1143	ANNEX	874	EVACUATE	1986
7	PENTAGON	1091	AUGMENT	871	SQUAD	1224
8	RIFLE	951	MODAL	688	LIEUTENANT	1213
9	PATROL	895	SENSOR	660	MILLILITER	1005
10	RADAR	720	LETHAL	659	RADIATE	993
사용범위 순위	7	빈도	8	빈도	9	빈도
1	DEPLOY	5254	ARMY	24471	THE	633238
2	TACTIC	4931	COMMAND	9223	AND	301202
3	HEADQUARTERS	3787	MILITARY	15403	BE	413347
4	SUBORDINATE	3212	SOLDIER	3138	OF	326223
5	CASUALTY	2041	MISSION	7750	TO	261252
6	GENERALS	1988	COMBAT	7811	IN	185133
7	ASSAULT	1694	ENEMY	7726	FOR	93837
8	BATTERY	1667	PERSONNEL	7139	THAT	123670
9	AVIATION	1589	STAFF	6463	OR	62456
10	SURVEILLANCE	1560	OFFICER	5643	OPERATE	36342

3.4.2 코퍼스 비교법

Chung과 Nation(2003)은 타겟 코퍼스(target corpus)와 비교 코퍼스들(comparison corpora)의 토큰 수가 같은 경우, 어떤 단어의 어휘 빈도수가 비교 코퍼스에 나온 해당 어휘 빈도수를 모두 더한 값과 비교해 50배 이상 차이가 나면 해당 단어의 어휘를 전문적인 어휘라고 보았다. 본 연구의 타겟 코퍼스 토큰 수는 약 1,000만 개이고, 비교를 위한 전문 코퍼스 8개의 총 토큰 수는 800만 개이다. 따라서 본 연구에서는 Chung과 Nation(2003)이 제시한 해당 비율 50:1을 40:1의 비율로 조정하여 적용하였는데, 그 결과 221개의 키워드가 도출되었으며 이는 비교 코퍼스들에

나타난 총 빈도수가 타겟 코퍼스에서 나타난 빈도수보다 40배 이상이 차이가 났다는 것을 의미한다.

표 4를 보면 Range 4 내 “Marine(3위)”과 “Synchronize(4위)”의 경우, 8개 비교 코퍼스에서 나타난 빈도수 합이 타겟 코퍼스 빈도수에 40을 곱한 수보다 크므로 표 5의 1:40 비율로 계산한 결과표에서는 제외된 것을 확인할 수 있다. 그러나 표 4에서와 마찬가지로, 표 5의 1:40의 비율이라는 기준을 적용한 결과 역시 어휘목록으로 그대로 사용하기에는 무리가 있어 보이는데, 이는 비율 적용으로 인해 목록에서 사라진 단어 어휘 중 일부도 연구자들은 교수·학습에 필요한 어휘라고 보았기 때문이다.

표 5. Range 범위별 1:40 비율적용 단어 리스트

사용범위 순위	1	빈도	2	빈도	3	빈도
1	INFANTRY	2808	BATTALION	5836	DECONTAMINATE	2090
2	TRUMP	944	PLATOON	3197	MUNITION	1491
3	HOWITZER	792	ARTILLERY	2646	MEDAL	1185
4	SNIPE	624	RECONNAISSANCE	2446	RANGER	886
5	ORDNANCE	576	MATERIEL	1173	RADIOLOGY	874
사용범위 순위	4	빈도	5	빈도	6	빈도
1	SERGEANT	5000	BRIGADE	5948	COM	4670
2	AMMUNITION	2237	TROOP	3057	CORPS	4522
3	ARMOUR	1226	FORT	2938	CIVILIAN	3672
4	HELICOPTER	1143	VETERAN	1350	LOGISTICS	2632
5	PENTAGON	1091	MAJORS	911	TERRAIN	2326
사용범위 순위	7	빈도	8	빈도	9	빈도
1	DEPLOY	5254	ARMY	24471	FIRE	13319
2	TACTIC	4931	COMMAND	9223	GUARD	4229
3	HEADQUARTERS	3787	MILITARY	15403		
4	CASUALTY	1988	SOLDIER	3138		
5	GENERALS	1694	MISSION	7750		

3.4.3 키워드 추출법

키워드는 코퍼스 언어학에서 많이 사용되는 개념으로서, 참조 코퍼스 대비 목표 코퍼스에서 현저히 높은 사용빈도를 보이는 어휘를 일컫는다(Baker 2004). Range 분석에서는 특정 분야별 코퍼스를 비교 코퍼스 파일로 사용했었지만, 이와 달리 AntConc 분석에서는 일반 코퍼스(general corpus)를 참조 코퍼스로 지정하였다. AntConc 프로그램은 해당 코퍼스 내 어휘와 참조 코퍼스 어휘를 타입(type)으로 비교한 후, 출현 빈도수가 높은 단어들과 비정상적으로 출현 빈도수가 낮은 단어들(negative keywords)을 카이제곱 검정 또는 로그 가능도 검정방식을 통해 중요도(keyness) 및 빈도수(frequency)를 산출하므로 단순 빈도수 활용을 통한 방식이 아닌, 통계적인 값에 의해 5,000개의 키워드를 추출한다. 아래 표 6은 AntConc 프로그램을 이용하여 keywords를 뽑아낸 결과를 정리한 것이다.

표 6. AntConc로 추출한 상위 키워드 결과표

순위	1	순위	2	순위	3
1	OPERATIONS	16	UNIT	31	COMMANDERS
2	ARMY	17	COMBAT	32	BRIGADE
3	SUPPORT	18	UNITS	33	TACTICAL
4	COMMANDER	19	FIRE	34	AIR
5	ATP	20	DEFENSE	35	Z
6	ENEMY	21	PLATOON	36	SUSTAINMENT
7	MILITARY	22	SOLDIERS	37	MEDICAL
8	MISSION	23	EQUIPMENT	38	CAPABILITIES
9	COMMAND	24	PLANNING	39	AREA
10	AND	25	BATTALION	40	STAFF
11	WKH	26	DQG	41	FIRES
12	FORCES	27	SECURITY	42	JOINT
13	FORCE	28	OPERATIONAL	43	WR
14	PERSONNEL	29	INFORMATION	44	THEATER
15	UNIT	30	TEAM	45	TERRAIN

표 6에서 볼 수 있듯이 AntConc 프로그램으로 keywords를 뽑아낸 결과, 상위 키워드 리스트 내에 “AND(순위10)”, “WKH(순위 11)”, “DQG(순위 26)”, “Z(순위 35)”, “WR(순위 43)”과 같이 키워드로 보이지 않는 단어들 또한 도출된 것을 확인할 수 있다. 이는 앞선 두 방법과 같이 세 번째 방법 역시 보조적인 용도로서 방대한 양의 리스트를 걸러주는 초벌 작업 정도의 구실은 할 수 있으나, 사람이 일정한 기준으로 소프트웨어적으로 분류된 어휘들을 재평가 및 분류해야 한다는 것을 의미한다.

4. 연구결과 및 분석

본 연구에서는 군사영어라는 특수한 분야의 코퍼스(MEA Corpus)를 구축하고, 이를 바탕으로 군사영어 어휘목록을 제시하고자 하였다. MEA Corpus를 만들기 위하여 사용된 총 토큰 수는 9,877,104개이고, 타입 수는 12,284개로서 표 7은 이를 정리한 것이다.

표 7. 구축된 코퍼스(MEA Corpus)에 쓰인 총 타입 및 토큰 수

Word List	Tokens / %	Types / %	Families
one	9,177,166 / 92.91	40,072 / 32.62	15,752
not in the lists	699,938 / 7.09	827,76 / 67.38	N/A
Total	9,877,104	122,848	15,752

4.1 어휘 선정 기준 적용

본 연구에서는 Range 분류법, 코퍼스 비교법, 키워드 추출법을 사용하여 어휘리스트를 만들고자 하였으며, 그 결과 군사영어 기초 어휘목록 개발에는 ‘Range 분류법’과 ‘코퍼스 비교법’을 활용하는 것이 가장 효과적이라고 보았다. 그러나 상기 방법으로 인한 단점을 보완하는 정교화 작업이

필요하다는 판단 아래, 본 연구에서는 군사영어 분야 전문가들의 expert judgment를 활용한 'familiarity'와 'specialty'의 어휘선정 기준 또한 적용하여 단어 리스트를 개발하였다. 이를 위하여 우선 군사영어를 가르친 경험이 있는 평가자 세 명의 친숙도 평가를 시행하였으며 그 결과, 표 5의 단어 15,747개 중 3,139개가 선정되었다.² 이렇게 선정된 3,139개의 단어에 군사영어 분야의 전문성을 바탕으로 “군사영어를 기준으로 삼은 친숙도”에 대한 점수가 부여되었다. 친숙도는 키워드 별 평가자의 평균점수 산출 이후, 1부터 9까지 각 Range에 해당하는 구간마다 평균점수 기준 내림차순으로 정렬하였다. 평가자 간 신뢰도(inter-rater reliability)의 Cronbach Alpha 값은 .877로 높은 편이었다.

또한, 3명의 외부 군사전문가들이 전문성(specialty)에 기초하여, 3,139개의 단어에 1에서 3까지의 점수를 부여하였다. 평가자 간 신뢰도 값은 .921이었으며, 세 명의 점수가 일치하지 않는 단어들은 협의를 거쳐 도출된 점수를 기반으로 3가지 카테고리(점수 1, 2, 3)에 따라 분류되었다. 전문성을 기준으로 평가된 단어 중 1점에 해당하는 단어들은 모두 제외시켰으며, 3점에 속하는 143개의 단어는 MEBWL_BASEWRD1로 명명하였다(표 8 참고). 또한, 2점에 속하는 3,602개의 단어 중 친숙도 기준 평균 2.67 이상인 단어 391개만을 따로 뽑아서 MEBWL_BASEWRD2로 이름 지었다(표 9 참고). 결과적으로, 이들 두 개의 리스트를 합친 어휘리스트인 MEBWL는 총 534개의 단어로 이루어져 있으며 표 8-9는 전문성 측정 척도 2 이상의 어휘로 구성된 MEBWL_BASEWRD1와 MEBWL_BASEWRD2 각각의 일부를 제시한 것이다.

표 8. MEBWL_BASEWRD1 어휘(전문성 측정 척도 3)

순위	어휘	순위	어휘	순위	어휘
1	ARMY	16	AMMUNITION	31	WARFARE
2	COMMAND	17	NAVY	32	MAJORS
3	MILITARY	18	MARINE	33	RANGER
4	SOLDIER	19	BATTLE	34	REGIMENT
5	COMBAT	20	GENERALS	35	ECHELON
6	BRIGADE	21	ASSAULT	36	HOWITZER
7	BATTALION	22	BATTERY	37	ENLIST
8	SERGEANT	23	AVIATION	38	MORTAR
9	PLATOON	24	MUNITION	39	SHRIMP
10	TROOP	25	SURVEILLANCE	40	CANNON
11	FORT	26	VETERAN	41	BREACH
12	INFANTRY	27	SQUAD	42	ORDNANCE
13	MANOEUVRE	28	LIEUTENANT	43	MEDIC
14	ARTILLERY	29	MISSILE	44	MINER
15	RECONNAISSANCE	30	TANK	45	CAVALRY

² 키워드의 빈도수가 20 이하인 경우, 활용도 측면에서 미미한 의미가 있을 것이라는 판단하에 평가에서 제외하였다.

표 9. MEBWL_BASEWRD2 어휘(전문성 측정 척도 2)

순위	어휘	순위	어휘	순위	어휘
1	OPERATE	16	DEPLOY	31	EFFECTIVE
2	UNIT	17	TACTIC	32	ISSUE
3	FIRE	18	ATTACK	33	LOGISTICS
4	MISSION	19	COM	34	AGENCY
5	DEFENCE	20	POSITION	35	RESERVE
6	DIVISION	21	COMPANY	36	TERRAIN
7	TEAM	22	VEHICLE	37	DECONTAMINATE
8	ENEMY	23	GUARD	38	AIRCRAFT
9	TASK	24	PREPARE	39	SUBORDINATE
10	TARGET	25	INTELLIGENCE	40	CASUALTY
11	PERSONAL	26	WEAPON	41	EVACUATE
12	STAFF	27	HEADQUARTERS	42	MARCH
13	BASE	28	CIVILIAN	43	OBJECTIVE
14	JOINT	29	THREAT	44	CONTACT
15	OFFICER	30	WAR	45	PHASE

4.2 MEABWL 어휘에 대한 타당성 분석

본 연구에서는 MEABWL 어휘의 타당성 검증을 위하여 텍스트 포괄범위를 분석하였다. 이를 위해 우선 AntWordProfiler_1.5.1을 활용하여 MEABWL의 BASEWRD1과 BASEWRD2를 BNC_COCA_25000의 첫 번째, 두 번째, 그리고 세 번째의 1000단어 수준의 리스트들을 함께 넣은 후 A1-E1의 텍스트에 이들 5개의 반영비율이 어떻게 나타나는지 알아보았다.

표 10. 군사 텍스트의 어휘 반영률 결과

	A1	B1	C1	D1	E1
MEABWL_BASEWRD1	0.4	0.3	0.6	0.2	0.5
MEABWL_BASEWRD2	1.0	0.9	1.1	1.0	1.2
BNC_1	59.4	60.2	59.8	62.1	60.5
BNC_2	17.8	16.3	15.6	15.7	16.0
BNC_3	12.1	12.4	11.2	10.7	12.2
Total	90.7%	90.1%	88.3%	89.7%	90.4%

위 표 8은 격년씩 발간되는 군사적인 맥락이 짙은 텍스트의 어휘 반영비율 분석결과이다. MEABWL_BASEWRD1의 단어 어휘목록은 전반적으로 0.3-0.5% 정도의 반영률을 나타냈으며, MEABWL_BASEWRD2의 단어 어휘목록의 반영률은 0.9에서 1.4% 정도의 반영률을 나타낸 것을 확인할 수 있다. 생성된 군사어휘 전문목록을 AntWordProfiler를 통해 실제 군사영어 수업현장에서 활용될만한 5개의 텍스트로 텍스트 포괄율(text coverage)을 확인한 결과, 군사영어 과목을 수강하는 수강생들이 한국의 영어 교과 교육과정을 충실히 수행해왔다는 가정하에 기본 BNC 1-3의 3000단어를 이미 숙지한 상황에서, 본 연구에서 도출해낸 MEABWL 어휘목록 내 총 534개의 단어를 추가적으로 학습한다면, 90%에 가까운 포괄율로 충분히 텍스트를 읽고 이해하는 데에 무리가 없을 것으로 보인다(이문복, 신동광 2015).

5. 논의 및 결론

본 연구의 목적은 다양한 군 관련 콘텐츠로부터 추출한 빅데이터 자료를 기반으로 구축한 군사 전문 코퍼스를 활용하여 어휘 목록을 개발, 제시함으로써 향후 군사영어 ESP 수업에서 학습자와 교수자 모두 효율적으로 활용할 수 있도록 도움을 주는 데에 있다. 이를 위하여 본 연구에서는 방대한 양의 구어 및 문어 텍스트를 수집하여 군사영어 전문 코퍼스를 구축하였다. 우선, 대규모 코퍼스를 바탕으로 군사영어에 기초가 되는 어휘목록을 구성하고자 하였는데, 이러한 작업에는 군사영어 맥락에서 필수 불가결한 어휘들을 정확하고 효율적으로 추출하는 기준과 방법이 필요하다. 따라서 연구자들은 4가지 선정 기준인 사용범위(range), 빈도수(frequency), 친숙성(familiarity), 전문성(specialty)을 염두에 두고 (1) 사용범위(range)에 기반해 8개의 ESP 비교 코퍼스를 활용하여 Range_Familizer25 프로그램으로 어휘를 추출한 후, (2) 빈도(frequency) 기준으로 효과적으로 어휘를 추출할 수 있는지 살폈으며, (3) 사람 평가자(human rater)의 전문성(expert judgment)을 바탕으로 친숙성(familiarity), 전문성(specialty)이라는 기준으로 어휘를 추려보았다. 이로써 본 논문은 다수의 ESP 어휘리스트 연구가 빈도 중심의 어휘목록을 제시한 데 비해 다양한 기준을 고려하여 효과적인 어휘 추출 방안을 탐색했다는 점에 있어 비슷한 성격의 후속 연구, 특히 군사영어 어휘목록 개발 연구를 위한 가이드 역할을 할 수 있을 것으로 기대되며, 또한 본 연구에서 사용한 Range 분류법, 코퍼스 비교법, 키워드 추출법이 어휘목록 개발에 있어 차별 작업으로써의 역할은 하지만, 전문가 판단을 바탕으로 추출된 어휘리스트와 비교해 중요 키워드들이 다수 누락된 것을 확인함으로써 어휘목록 개발에 있어 컴퓨터 프로그램이 아직은 ‘보조적’ 수단으로서의 역할을 할 수 밖에 없음을 시사한다.

그러나, 본 연구는 한정된 시간 내에 상이한 방식으로 텍스트를 수집하다 보니 코퍼스의 베이스가 된 문어와 구어의 비율을 균등하게 통제하지 못했다는 한계점을 가진다. 또한, STT 기술을 활용하여 자동생성되는 유튜브 동영상 자막 추출 과정을 통해 대다수의 구어 자료를 얻을 수 있었으나, 현행 STT 기술의 불완전성으로 인해 수작업으로 텍스트를 사후 보정해야 하는 번거로움 또한 제한점이라 할 수 있겠다. 그럼에도 불구하고 문어 데이터 중심으로 편중되어 있던 기존 선행 연구들의 코퍼스와 달리 최신의 기술을 활용해 수집한 구어와 문어 데이터를 바탕으로 특수목적영어 전문 코퍼스를 구축을 했다는 점에 있어 의의가 있다.

또한, 의학영어, 호텔관광영어, 컨벤션 영어 등 해당 분야의 ESP 어휘리스트를 빈도수 기준으로 개발하여 다룬 대다수의 기존 연구와 달리 특수목적영어의 특성을 고려해 전문성, 친숙성, 사용범위, 빈도를 기준으로 군사영어 전문 어휘목록을 개발하여 제시한 최초의 시도였다는 점과 빅데이터 코퍼스를 활용하여 특수목적 어휘 목록을 개발함으로써 이러한 접근법이 군사영어를 넘어 의학영어, 항공영어, 미용영어 등 다양한 분야로 외연을 확장할 수 있는 가능성을 보여주었다는 점에서 본 연구는 시사점을 가진다.

군사영어 및 ESP 어휘교육에 대해 관계자들이 가지는 관심 및 필요성에 비해 관련 ESP 연구가 현저히 부족할 뿐 아니라 군사영어 어휘목록이 아직 개발되지 않은 실정이다. 이러한 상황을 고려해 볼 때, 본 연구는 빅데이터 활용 전문 코퍼스 구축 및 이를 기반으로 한 특수목적영어 어휘리스트를 개발, 제시하고 ESP 수업에 활용할 수 있는 방안을 제안했다는 점에서 학문적 의의를 찾아볼 수 있다.

참고 문헌

- 강남희(Kang, N.). 2017. 플립러닝을 활용한 간호영어(ESP) 수업 사례 연구(A case study of ESP nursing course based on flipped learning). 《문화와 융합》 (*Culture and Convergence*) 39-5, 323-336.
- 강범모(Kang, B.). 2003. 언어, 컴퓨터, 코퍼스 언어학(*Language, Computer, Corpus Linguistics*). 서울: 고려대학교 출판부(Seoul: Korea University Press).
- 권선희(Kwon, S.). 2013. 컨벤션영어 텍스트의 어휘 목록 생성과 분석(Profiling the vocabulary lists in a convention English corpus). 《영어교육연구》 (*English Language Teaching*) 25-4, 145-165.
- 권은영(Kwon, E.-Y.). 2020. 파이썬과 텍스트 마이닝을 활용한 ESP 연구 동향(A study of research trends in ESP using python and text mining). 《외국어교육》 (*Foreign Languages Education*) 27-2, 111-139.
- 김나래·이한솔·정태영(Kim, N., H. Lee and T.-Y. Jeong). 2019. 특수목적영어 학습 및 평가에 미치는 학습동기와 학습전략에 관한 연구: 군사영어 학습을 중심으로(Motivation and strategies in ESP learning and assessment: A case of military English in South Korea). 《영어평가》 (*English Language Assessment*) 14-1, 109-132.
- 김미영(Kim, M.). 2019. 구어 항공관제영어 코퍼스 기반 어휘 연구(A Corpus-based lexical analysis of spoken air traffic control English). 《코기토》 (*Kogito*) 89, 349-376.
- 김성식(Kim, S.). 2007. 코퍼스와 영어교육(Corpus and English education). 《영어교육》 (*English Teaching*), 62-2, 281-307.
- 김태국·정채관(Kim, T. and C. Jung). 2016. 교육적 코퍼스 활용에 관한 학습자의 태도 분석: 고등학생을 중심으로(Korean high school student attitudes toward pedagogic corpus use). 《중등영어교육》 (*Secondary English Education*) 9-1, 47-80.
- 송근호(Song, G.). 2022. 한국군 군사영어 소통능력 향상을 위한 영어교재 개선방안 연구(A study on the improvement of ELT materials to enhance the military English communication skills of the Korean armed forces). 《한국군사학논집》 (*Korean Journal of Military Art and Science*) 78-1, 349-374.
- 신경식·박지환(Shin, K. and J. Park). 2009. 『코퍼스 분석을 통한 군사영어 어휘 연구』 (*A Study on Military English Vocabulary Using Corpus Analysis*). 서울: 화랑대연구소 (Seoul: Hwarangdae Research Institute).
- 신동광(Shin, D.). 2011. 기본 어휘의 선정 기준: 영어 어휘를 중심으로(Criteria for selecting basic words: Focusing on English vocabulary). 《국어교육학연구》 (*Korean language education research*) 40, 217-243.
- 신동광·주현우(Shin, D. and H. Joo). 2008. 영어과 교과서 검정용 어휘 목록 개발: 어휘 검색 프로그램 및 기본 어휘 목록에 대한 고찰(Development of a vocabulary list for English textbook authorization: Reflection on vocabulary analysis programs and the basic vocabulary list). 《멀티미디어 언어교육》 (*Multimedia-Assisted Language Learning*)

- 11-3, 93-111.
- 신창원(Shin, C.). 2012. ESP 교육을 위한 소규모 코퍼스 구축: 기계공학 관련 논문초록을 중심으로(The compilation of a small corpus for ESP teaching: A special corpus of research paper abstracts in mechanical engineering domain). 《언어와 정보 사회》 (*Language & Information Society*) 17, 175-205.
- 이경은·신정화(Lee, K. and J. Shin). 2017. 한국인 영어 학습자의 관계대명사 사용에 관한 코퍼스 기반 연구(A corpus-based study on the use of relative pronouns by Korean learners of English). 《영어학》 (*Korean Journal of English Language and Linguistics*) 17-4, 817-837.
- 이문복·신동광(Lee, M. and D. Shin). 2015. 2015 영어과 교육과정 기본 어휘 목록 개발 (Development of the Korean basic English word list of the 2015 revised national curriculum of English). 《영어교과교육》 (*Korean Journal of English Language and Linguistics*) 17-4, 817-837.
- 이성은(Lee, S.). 2022. 의학영어 어휘목록 개발: 코퍼스 기반 연구(Developing a word list in medicine: A corpus-based study). 《인문사회21》 (*The Journal of Humanities and Social Science*) 13-4, 2739-2752.
- 이제영·김정렬(Lee, J. and J. Kim). 2013. 학술 전문 코퍼스를 활용한 영어교육 전문 어휘 추출 (Extracting technical vocabulary list for English Education: A corpus-based research). 《외국어교육》 (*Foreign Languages Education*) 20-1, 99-121.
- 이한솔·신경식·이장호(Lee, H., K. Shin and J. Lee). 2012a. 실용자료에 근거한 군사영어 코퍼스에 관한 연구(A Study on military English corpus based on authentic materials). 《한국군사학논집》 (*Korean Journal of Military Art and Science*), 68-1, 239-263.
- 이한솔·이장호·신경식(Lee, H., J. Lee and K. Shin). 2012b. 코퍼스 기반 연구의 한미 연합작전 활용방안-연합작전 임무 수행에 관련된 군사어휘 연구(How a corpus-based study can help in ROK-US combined operations: A corpus-based study on the military vocabulary). 《한국군사학논집》 (*Korean Journal of Military Art and Science*) 68-3, 49-72.
- 이한솔·김나래(Lee, H. and N. Kim). 2020. 특수목적영어 학습전략, 동기, 성취의 관계 연구: 군사영어를 중심으로(Exploring the relationship among language learning strategies, motivation, and achievement in ESP learning: A case of learning military English). 《한국군사학논집》 (*Korean Journal of Military Art and Science*) 76-2, 283-310.
- 이한솔·오준혁·정태영(Lee, H., J. Oh and T.-Y. Jeong). 2021. 특수목적영어 학습동기 향상을 위한 형성평가 활용 사례연구: 군사영어 학습을 중심으로(Implementation of formative assessment for motivating ESP learning: A case study of military English learning). 《한국군사학논집》 (*Korean Journal of Military Art and Science*) 77-3, 281-305.
- 이효진·민덕기(Lee, H. and D. Min). 2016. 초등학교 영어교육에서 코퍼스 분석을 통한 전치사 교육(A study on uses of prepositions in English corpora for elementary school students.) 《초등영어교육》 (*Primary English Education*). 22-1, 61-86.

- 임수영·이은주(Im, S. and E. Lee). 2012. 코퍼스를 활용한 문법 과제가 고등학생의 영어 문법 학습과 정의적 반응에 미치는 영향(The effects of the corpus-based grammar tasks on high school students' English grammar learning and attitude). 《영어학》 (*Journal of English Language and Linguistics*) 12-2, 303-325.
- 장세은·이성민(Jang, S. and S. Lee). 2013. Visualization of collocational networks: Maritime English keywords. 《어학연구》 (*Language Research*) 49-3, 781-802.
- 장철원(Jang, C.). 2022. 『웹 크롤링 & 데이터 분석 with 파이썬』 (*Web Crawling & Data Analysis with Python*). 서울: 인사이트(Seoul: Insight).
- 정미애(Jung, M.), 2007. 비즈니스 영어 코퍼스 생성과 분석(Building and analyzing a corpus of business English). 《언어연구》 (*Journal of Studies in Language*) 23-3. 429-451.
- 정채관(Jung, C.). 2021. 한국인 영어 학습자 구어 코퍼스: 설계와 구축(Designing and building the Korean English Learners' Spoken Corpus). 《외국어교육연구》 (*Studies in foreign language education*) 35-3. 209-223.
- 정태영(Jeong, T.-Y.). 2018. K-COPI 기반 학습자 구어 영어 코퍼스 구축 및 활용(Building and using K-COPI based learners spoken English corpus). 《언어와 정보 사회》 (*Language & Information Society*) 33. 251-274.
- 정태영·신경식·허진(Jeong, T.-Y., K. Shin and J, Heo). 2009. 『Corpus를 이용한 ESL 학생의 영어 말하기 오류 분석』 (*Corpus-based error analysis in ESL students' English oral performance*). 서울: 화랑대연구소(Seoul: Hwarangdae Research Institute).
- 진성환(Jin, S.). 2022. 육군사관학교 영어교육을 중심으로 본 군사영어교육 발전방안: 국제언어로서의 영어교육 적용의 필요성(TEIL in the military context: Exploring the possibilities for the Korea Military Academy). 《한국군사학논집》 (*Korean Journal of Military Art and Science*) 78-2, 55-86.
- 최인영(Choi, I.), 2020. 코퍼스(Corpus) 기반 항공안전영어 어휘 리스트 구축: IATA의 Safety Report를 중심으로(A corpus-based study on the keyword of airline safety English: Focused on safety report published by IATA). 《관광레저연구》 (*Journal of Tourism and Leisure Research*) 32-12, 299-316.
- 한나래·이수화(Han, N.-R. and S.-H. Lee). 2009. 학습자 코퍼스를 이용한 영어 전치사 오류 교정 모델 개발(Developing a model for English preposition errors using a learner corpus). 《언어학》 (*Journal of the Linguistic Society of Korea*) 53, 163-185.
- 황필아(Hwang, P.). 2012. 학문적 목적의 구두/청취 영어에 대한 과업기반 요구분석: 한국과 중국 유학생을 중심으로(Task-based needs analysis of academic oral/aural skills of Korean/Chinese international students in the U.S. universities). 《영어학》 (*Journal of English Language and Linguistics*) 12-1, 149-178.
- Baker, P. 2004. Querying keywords: Questions of difference, frequency and sense in keywords analysis. *Journal of English Linguistics* 2-4, 346-359.
- Biber, D., S. Conrad, and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

- Carter, R. 2012. *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.
- Chung, T. M. and I. S. P. Nation. 2003. Technical vocabulary in specialized texts. *Reading in a Foreign Language* 15-2, 103-116.
- Coxhead, A. (1998). The development and evaluation of an academic word list. Unpublished master's thesis. Wellington: Victoria University of Wellington
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34, 213-238.
- Coxhead, A. 2013.
- Cummins, J. 1979. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research* 49, 222-251.
- Dudley-Evans, T. and M. St. John. 1998. *Developments in ESP: A Multi-Disciplinary Approach*. Cambridge: Cambridge University Press.
- Jhang, S.-E. and K. Parent. 2011. The vocabulary of maritime English: Keyword analyses of the English homepages of port authorities around the world. *Journal of English Language and Linguistics* 11-4, 1065-1083.
- Johns, A. and T. Dudley-Evans. 1991. English for specific purposes: International in scope, specific in purpose. *TESOL Quarterly* 25, 297-314.
- Kim, D. 2015. A profile and characteristics of military English. *Journal of Humanities* 19, 259-290.
- Lei, L. and D. Liu. 2016. A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes* 22, 42-53.
- Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. and K. Hwang. 1995. Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23, 35-41.
- Noguera-Díaz, Y. and P. Pérez-Paredes, 2019. Register analysis and ESP pedagogy: Noun-phrase modification in a corpus of English for military navy submariners. *English for Specific Purposes* 53, 118-130.
- Orna-Montesinos, C. 2013. English as an international language in the military: A study of attitudes. *ESP Journal* 4-1, 87-105.
- Pierce, L. and M. Dixon. 2006. *Improving Multicultural Teamwork to Combat Terrorism*. (ADA474095). Army research lab Aberdeen proving ground MD human research and engineering directorate. <https://apps.dtic.mil/sti/pdfs/ADA474095.pdf>
- Valipouri, L. and H. Nassaji. 2013, A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes* 12-4, 248-263.
- Wang, J., S. L. Liang, and G. C. Ge. 2008. Establishment of a medical academic word list. *English for Specific Purposes* 27-4, 442-458.
- Ward, J. 1999. How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language* 12-2, 309-323.

West, M. 1953. *A general Service of English Words*. London: Longman.

Wilkins, D. 1972. *Linguistics in Language Teaching*. London: Arnold.

Xue, G. and I. S. P. Nation. 1984. A university word list. *Language Learning and Communication* 3-2, 215-229.

Yang, M. 2015. A nursing academic word list. *English for Specific Purposes* 37, 27-38.

예시 언어(Examples in): English

적용가능 언어(Applicable Languages): English

적용가능 수준(Applicable Level): Tertiary