

## 컴퓨터 공학 분야 영어 학술 논문의 전문 어휘 분석: 코퍼스 기반 연구

이제영\* · 이혜진\*\*

전주대학교

원광대학교

### Technical Vocabulary in Computer Science Research Articles: A Corpus-based Study

Lee, Je-Young\* and Lee, Hye Jin\*\*

Jeonju University

Wonkwang University

\*First Author / \*\*Corresponding Author

 OPEN ACCESS



<https://doi.org/10.18627/jslg.36.4.202102.487>

pISSN : 1225-4770

eISSN : 2671-6151

**Received:** January 06, 2021

**Revised:** February 01, 2021

**Accepted:** February 08, 2021

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

Copyright©2021 the Modern Linguistic Society of Korea

본인이 투고한 논문은 다른 학술지에 게재된 적이 없으며 타인의 논문을 표절하지 않았음을 서약합니다. 추후 중복게재 혹은 표절된 것으로 밝혀질 시에는 논문게재 취소와 일정 기간 논문 제출의 제한 조치를 받게 됨을 인지하고 있습니다.

#### ABSTRACT

*The Journal of Studies in Language* 36.4, 487-497. Computer science is a fast-moving and must-learn discipline, with a high lexical density of conceptually-demanding technical vocabulary. This study seeks to identify technical vocabulary used within the discipline of computer science through the corpus-based methodology. A specialized corpus of 135 research articles in computer science was built, comprised of 1,260,460 running words and for the analysis, AntWordProfiler was utilized. The corpus-driven analysis demonstrated that the GSL accounted for 71%. GSL and AWL's combination achieved 83.1% of the computer science corpus, less than the 95% lexical threshold suggested for optimal comprehension. These results highlight the need for a mastery of technical lexis in computer science, which was found to be technical; a total of 545 field-oriented technical words were identified. (Jeonju University · Wonkwang University)

**Keywords:** computer science, technical vocabulary, specialized corpus, discipline-specific, corpus-driven analysis

#### 1. 서론

본 연구는 컴퓨터 공학을 위한 전문어휘 목록이 개발되어야 할 필요성에 따라 시행되었다. 최근 기술시장 분석조사 전문기관인 Gartner는 2020년 10대 기술전략 중 SW(소프트웨어) 교육이 주목해야 할 것으로 전문성의 민주화(democratization of expertise)를 내세웠다. 즉, 과거의 소수 SW 전문가 영역으로 간주되었던 SW 제작 영역(예, 챗봇 제작, 머신러닝, 어플 개발)이 자동화 되면서 로코드(low-code) 및

노코드(no-code)로 인해 전문교육을 받지 않은 일반인도 쉽게 접근할 수 있게 되었다는 것이다. 과거에는 전문가가 아니면 접근하기 어려웠던 첨단 기술들이 일반인에게 대중화되었고 이에 따라 컴퓨터공학의 교육적 수요 역시 나날이 증대되고 있다. 문제는 초중고 및 대학생을 넘어 일반인까지 컴퓨터 교육이 강조되는 강도에 비해 컴퓨터 용어 관련 교육은 상대적으로 미진하다는 것이다.

일상생활에서 컴퓨터 사용은 필수가 되었고, 컴퓨터 시스템 학습자, 관련 업계 종사자뿐만 아니라 일반인들이 컴퓨터 관련 텍스트를 읽을 때 마주하게 되는 장애물 중 하나는 어휘일 것이다. 컴퓨터공학은 특히 학문의 변화 속도가 빠르며, 새로 생성되는 대부분 정보가 영어로 가장 먼저 그리고 많이 만들어진다는 현실을 고려해보면 국제 경쟁력을 갖추는 차원에서 컴퓨터공학 영어어휘는 반드시 습득해야 한다. 특히 컴퓨터 공학과 같이 전문성이 강한 학문 분야에서 전문어휘에 대한 논의는 더 적극적이고 직접적으로 이뤄져야만 한다.

특히 컴퓨터 공학 분야의 학부생 및 대학원생들의 경우 대학과 같은 고등교육기관에서 영어 원서 혹은 학술지에 실린 논문을 통해 학문 목적 영어(English for Academic Purposes, 이하 EAP)로서 컴퓨터 전공 내용을 자주 접하게 된다. 그러나 대부분 수업에서 전공과목의 내용(contents)에만 치중할 뿐 실제 원서나 영어문서에서 사용되는 어휘, 즉 EAP 교육이 직접적으로 이뤄지는 경우는 많지 않다(이제영·김정렬, 2013). 컴퓨터 공학 분야의 EAP의 경우 영어로 말하거나 발표하는 경우보다 영어 문서를 읽고 이해해야 하는 경우가 더 많은데 이때 어휘 지식이 핵심 역할을 한다. 많은 학자들(이제영·김정렬, 2013; Coxhead and Demecheleer, 2018; Masrai and Milton, 2018)은 EAP 학습에 대한 부담을 경감시키고 효율성을 극대화시키기 위해서는 명시적인 어휘 교육이 필요하며, 특히 의학과 공학과 같이 전문성이 강한 학문 분야에서는 전문어휘에 대한 논의가 더 적극적이고 직접적으로 이뤄져야 한다고 주장하였다.

이러한 교육적 필요성에 입각하여 본고는 컴퓨터공학 분야의 논문에서 주요 핵심 전문어휘를 추출하고자 한다. 논문은 전공 서적이거나 일반도서와는 달리 저자가 새롭거나 창의적인 아이디어가 떠오르면 빠르게 게재할 수 있기 때문에 최신 방법론 및 관련 기술의 발전 동향, 최근 이슈가 되는 쟁점, 학계의 관심 분야 등을 더욱 빠르게 습득할 수 있다. 이와는 반대로 책은 저자가 출판사와 계약을 맺고 편집, 인쇄, 검수 및 검증 절차 등의 일련의 작업을 요구하기 때문에 최신 경향에 대한 반영이 늦어질 수 있다. 특히 컴퓨터공학 분야와 같이 신속하게 발전되는 분야일수록 전문용어의 통용성 제고를 위해서는 논문이 적합하다고 판단되어 본 연구는 컴퓨터공학에서 사용되는 전문어휘를 목록화하고자 한다.

전문어휘란 학문의 분화 혹은 전문화로 인하여 새롭게 생겨난 지식이나 또는 개념 등을 표현하기 위하여 만들어지거나 도입된 용어이다(설문원, 2003). 전문어휘는 주로 해당분야 혹은 특정 집단에서 정교하게 규정된 학술적 개념을 포함한 특수한 어휘로 해당 전공 학생의 전문지식 향상 및 학습에 도움을 줄 수 있다. 전문어휘를 추출하는 방법에는 여러 가지가 있으나 본 연구에서는 코퍼스를 이용하여 컴퓨터 공학이라는 특수 학문 분야에서 주로 사용되는 실제 언어를 분석하고자 한다. 코퍼스란 방대한 규모의 언어 데이터 자료를 분석할 수 있도록 전산화한 것으로 체계적인 원칙을 따라 구축하였으므로 대표성(representativeness)을 지닌다. 코퍼스 언어학의 발전으로 국내 많은 선행연구에서 코퍼스 분석을 통한 전문어휘 추출을 시행하였으며 학문분야에 따라 유아교육(이제영 등, 2017), 해사(장세은·변현정, 2011), 공학(하명호, 2017), 영어교육(이제영·김정렬, 2013; 하명호, 2018), 음성학 및 음운론(이제영·이혜진, 2019) 등 제반 영역에서 이루어졌음을 알 수 있다. 위 선행 연구들 모두 각 학문 분야에서 학술적 및 교육적 공헌을 하였지만 정작 컴퓨터 교육이 점차적으로 확대 및 심화되는 현시점에서 컴퓨터 공학 전문어휘 관련된 연구는 미진한 것으로 사료된다. 이에 본 연구는 컴퓨터공학 관련 교육에 실질적인 공헌을 하고자 다음과 같은 연구 질문을 탐구하고자 한다.

첫째, 컴퓨터공학 코퍼스에서 추출된 일반 어휘(GSL) 및 학술어휘(AWL)의 분포 양상은 어떠한가?  
 둘째, 컴퓨터공학 코퍼스에 자주 등장하는 전문어휘(TV)는 어떠한 것들이 있는가?

## 2. 선행 연구

### 2.1 어휘의 종류

Coxhead와 Nation(2001)은 학술텍스트에 등장하는 어휘를 전문성을 토대로 기본어휘(basic vocabulary), 학술어휘(academic vocabulary), 전문어휘(technical vocabulary), 저빈도 어휘(low-frequency vocabulary)로 구분하였다. 여기서 전문성이란 해당 어휘의 사용이 특정 학문 분야에 제한되는 정도를 의미하는 것으로 전문어휘는 특정 학문에만 나타나거나 일상적인 의미와는 다른 의미로 사용되는 어휘로 간주될 수 있다. 아래 <그림 1>은 Coxhead와 Nation(2001: 252)이 구분한 4가지 어휘의 종류를 도식화하여 제시한 것이다.

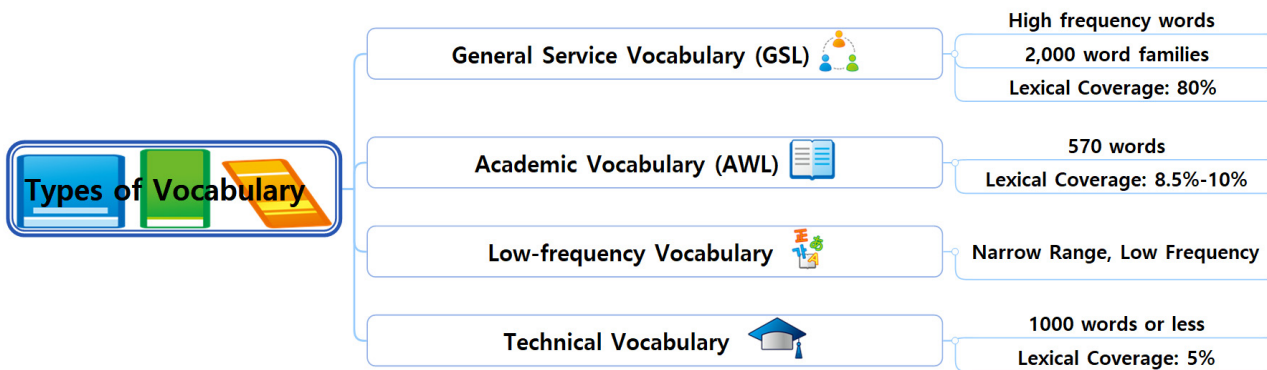


그림 1. 어휘의 종류(Coxhead and Nation, 2001: 252)

위에 구분한 어휘분류를 토대로 특정 학문을 위한 어휘 교육적 측면에서 논하자면 다음과 같다. 첫째, 기본어휘란 특정 텍스트에 80% 가량 등장하는 어휘로 일반적으로 기본적인 의사소통을 위해 사용되는 어휘이다(예: *the, and, get, have* 등). 텍스트에 등장 비율이 높고 학습 초기에 교육이 이뤄졌을 가능성이 높으며 일상생활에서 쉽게 접근 및 습득될 수 있다는 점에서 특정 학문을 위해 별도로 교육해야 할 필요성이 적다. 둘째, 학술어휘는 여러 학문 분야의 학술 텍스트에서 범용적으로 사용되는 어휘이며, 학술 텍스트의 약 10% 정도 차지한다고 알려져 있다(예: *analyze, define, formula, involve* 등). 학술어휘는 여러 학문 분야에 두루 나타나기 때문에 컴퓨터 공학을 비롯한 다양한 학문 분야를 영어로 공부해야 하는 학생들이 자신의 학술 목적 영어 능력을 키우기 위해 꼭 필요한 어휘이다. 셋째, 저빈도 어휘는 문자 그대로 출현빈도가 낮은 어휘이므로 별도의 시간을 할애하여 교육 또는 학습할 필요성이 낮다. 마지막으로 전문어휘는 통상 텍스트의 5% 정도의 등장 비율을 갖는 어휘로 알려져 있으며 근원적으로 해당 분야의 특수성 및 전문지식을 표상하는 특징이 있다.

학술 텍스트를 사전의 도움 없이 원활히 읽고 이해하기 위해서는 해당 텍스트에 등장하는 어휘의 최소 95% 이상을 파악하고 있어야 한다(Hirsh and Nation, 1992). Coxhead와 Nation(2001)에 의하면 학술 텍스트의 약 8.5%-10% 비율을 점유하고 있는 학술어휘(AWL, 570단어)와 약 80% 정도를 차지하는 일반어휘(GSL, 2000단어)를 습득하면 학술텍스트에 등장하는 어휘의 약 90%가 습득이 가능하며 여기에 5% 정도 차지하고 있는 전문어휘가 추가된다면 이해의 범위는 약 95% 정도의 산술적 기준을 충족하게 되므로 해당 분야의 학술텍스트를 읽는데 큰 어려움이 없어진다고 한다.

## 2.2 전문어휘 및 선행연구

전문용어는 해당 학문 분야의 내용을 습득하거나 정확한 정보를 습득하기 위해서 선제적으로 갖춰야 할 핵심요소로 일상에서 전반적으로 사용하는 언어와 다르게 전문분야에 종사하거나 관련 분야에 종사하는 사람들에 의해 구축되고 사용되는 용어를 의미한다(하명호, 2017). Chung과 Nation(2004)은 전문용어를 추출하는 방법으로 전문용어사전 사용, 전문가 기반 평가척도 사용, 텍스트에 제공된 단서 활용, 코퍼스 접근법을 제시하였다. 해당 분야 전문가들을 중심으로 전문 어휘를 선별한다면 비용과 시간이 많이 소요될 뿐만 아니라 전문가 구성에 따라 일관성을 유지하기가 어려우므로 제약이 따른다. 전문용어사전은 언어의 통용성을 반영하기가 어려우며, 텍스트에 제공된 단서를 통해 전문어휘를 추출하는 방식 역시 많은 인력과 시간이 요구되기에 현실적 어려움이 따른다. 이를 해결한 방법으로 국내외 많은 연구에서 최근 대두되고 있는 자연어 처리를 통한 코퍼스 분석을 시행하였으며, 앞서 서론에서 언급한 것처럼 여러 학문 분야에서 전문 어휘 추출이 이루어지고 있다.

### 학문간 인접정도에 따른 학문 분류

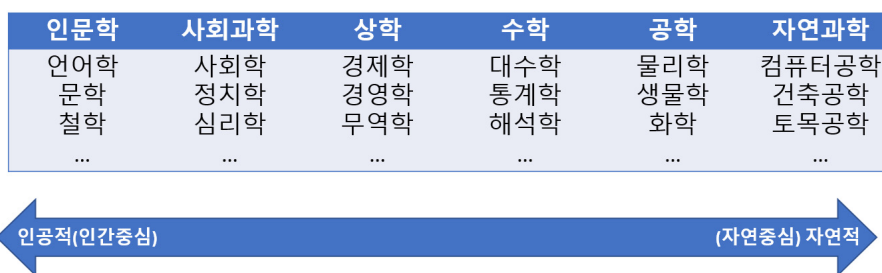


그림 2. 학문간 인접정도에 따른 분류(황성은, 2018: 358)

코퍼스를 이용하여 공학(기계, 화학, 전기, 컴퓨터, 환경, 전자, 산업, 재료) 분야의 공통 어휘목록을 개발한 하명호(2017)는 GSL(680개), AWL(233개), TWL(357개)로 구성된 총 1,170개의 공학 어휘목록을 제시하였으며, 유아교육 분야에서는 이제영 등(2017)이 GSL(2,000개), AWL(570개), TWL(224개)의 어휘목록을, 영어교육 분야에서 하명호(2018)는 GSL(457개), AWL(215개), TWL(72개)를 포함한 총 744개의 어휘 목록을, 음성학/음운론 분야에서 이제영과 이해진(2019)는 빈출 학술어휘와 372개의 빈출 전문어휘 목록을 제시하였다. 이처럼 해당 코퍼스가 가지고 있는 사용역(register), 장르(genre), 학문 분야(field-specific) 등의 요인에 따라 특수한 어휘 사용양상을 보여준다는 것이다. 특히, 황

성은(2018)이 인접 정도에 따라 구분한 학문분류(<그림 2>)를 토대로 보면 학문간 거리가 멀수록 언어의 이질성이 높다. 이는 영어에 능통한 영어영문학 전공자라도 컴퓨터 관련 영어 텍스트를 읽을 때 어려움이 있을 수 있다는 것이다. 그럼에도 불구하고 기존 선행 연구 중 기계, 화학, 전기, 컴퓨터 등 공학 분야 전반에 등장하는 전문 어휘를 분석한 하명호(2017)의 연구가 있을 뿐, 컴퓨터 공학 분야의 전문 어휘를 집중적으로 분석한 연구는 부족했다. 전공을 막론하고 컴퓨터에 대한 이해는 필수인 시대이므로 이에 본 연구는 컴퓨터 공학 분야에서 자주 등장하는 전문어휘를 도출하여 발전에 이바지하고자 한다.

### 3. 연구 방법

#### 3.1 분석대상

본 연구에서 활용한 컴퓨터 공학 코퍼스(Computer Science Corpus, 이하 CSC)는 ACM (Association of Computing Machinery)에서 출간하는 학술지에 수록된 학술 논문을 기초로 구축하였다. ACM은 IEEE Computing Society와 더불어 컴퓨터 과학 분야의 국제 저명 학술지를 발간하는 기관으로 널리 알려져 있으며, 컴퓨터 시스템, 하드웨어, 소프트웨어, 네트워크, 보안, 컴퓨터 교육 등 컴퓨터 공학의 여러 세부 분야와 관련된 다양한 학술지를 출간하고 있다. 또한 ACM에서 출간하는 학술지들은 한국교육학술정보원에서 운영하는 RISS의 해외전자자료검색 기능을 통해 무료로 접근할 수 있다는 편의성도 갖추고 있다.

구축하는 코퍼스에 다양한 컴퓨터 공학의 세부 분야를 다룬 학술 논문이 포함되도록 하기 위해 다음과 같은 절차를 통해 코퍼스 구축을 진행하였다.

- (1) ACM에서 출간하는 학술지 중 SCIE에 등재된 총 27종에 학술지(<표 1>)를 선정하였다.
- (2) 2016~2018년의 3년간 해당 학술지에 수록된 논문을 학술 데이터베이스를 통해 PDF 파일 형태로 수집하였다.
- (3) 각 학술지 당 5편의 논문(총 135편)을 임의로 선택하고, 선택된 학술 논문을 TXT 형태로 변환하였다.

**표 1.** CSC에 포함된 ACM 학술지 목록(알파벳순)

순번	학술지명
1	ACM Transactions on Algorithms (TALG)
2	ACM Transactions on Applied Perception (TAP)
3	ACM Transactions on Architecture and Code Optimization (TACO)
4	ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)
5	ACM Transactions on Autonomous and Adaptive Systems (TAAS)
6	ACM Transactions on Computer Logic (TOCL)
7	ACM Transactions on Computer Systems (TOCS)
8	ACM Transactions on Computer-Human Interaction (TOCHI)
9	ACM Transactions on Computing Education (TOCE)
10	ACM Transactions on Database Systems (TODS)
11	ACM Transactions on Design Automation of Electronic Systems (TODAES)

표 1. CSC에 포함된 ACM 학술지 목록(알파벳순) (Continue)

순번	학술지명
12	ACM Transactions on Embedded Computing Systems (TECS)
13	ACM Transactions on Graphics (TOG)
14	ACM Transactions on Information Systems (TOIS)
15	ACM Transactions on Intelligent Systems and Technology (TIST)
16	ACM Transactions on Internet Technology (TOIT)
17	ACM Transactions on Knowledge Discovery from Data (TKDD)
18	ACM Transactions on Mathematical Software (TOMS)
19	ACM Transactions on Modeling and Computer Simulation (TOMACS)
20	ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)
21	ACM Transactions on Privacy and Security (TOPS)
22	ACM Transactions on Programming Languages and Systems (TOPLAS)
23	ACM Transactions on Reconfigurable Technology and Systems (TRETSS)
24	ACM Transactions on Sensor Networks (TOSN)
25	ACM Transactions on Software Engineering and Methodology (TOSEM)
26	ACM Transactions on Storage (TOS)
27	ACM Transactions on the Web (TWEB)

위와 같은 절차를 통해 구축된 컴퓨터 공학 코퍼스는 126만 단어 수준이었다. 본 연구와 유사한 학술 영어 전문 코퍼스들의 규모를 살펴보면, Chen과 Ge(2007)가 구축한 Whole Paper Corpus는 190만 단어 수준, Vongpumivitch et al.(2009)의 The Applied Linguistics Research Article Corpus는 150만 단어 수준, Martínez et al.(2009)의 AgroCorpus는 80만 단어 수준으로 본 연구의 컴퓨터 공학 코퍼스와 비슷한 수준이었다.

### 3.2 분석 도구 및 기준

본 연구에서 전문 어휘를 추출하는데 사용된 도구는 Anthony(2014)가 개발한 무료 어휘 분석 소프트웨어인 AntWordProfiler 1.4.1이다. 이 프로그램은 GSL(General Service List; West, 1953)과 AWL(Academic Word List; Coxhead, 2000)과 같은 기존 개발된 어휘 목록, 또는 연구자가 특정 목적을 위해 작성한 어휘 목록에 포함된 어휘의 빈도와 분포를 분석하는 기능을 제공한다.

전문 어휘의 선정은 보통 특정 어휘의 출현 빈도(frequency)와 분포(range or distribution)를 기준으로 이루어지며, 본 연구에서도 이 두 가지 기준을 활용하였다. 먼저 빈도는 Coxhead(2000)가 350만 단어에서 100회 출현한 어휘를 학술 어휘로 선정했던 기준을 활용하여 126만 단어 수준인 본 연구에서는 36회 이상을 빈도에 대한 기준으로 설정하였다. 분포에 대한 기준은 특정 단어가 특정 텍스트에서만 여러 번 등장하는 경우를 방지하기 위해 설정되며, 본 연구에서는 총 135개 논문의 5% 수준인 최소 7편 이상의 논문에 등장한 어휘를 전문 어휘에 포함시켰다.

## 4. 연구 결과 및 논의

### 4.1 어휘의 분포

먼저 기본 어휘인 GSL과 학술 어휘인 AWL의 분포를 분석하여 아래 <표 2>에 제시하였다.

표 2. 기본 어휘와 학술 어휘의 분포

GSL 1,000	GSL 2,000	AWL	기타	합계
833,508 (66.1%)	61,825 (4.9%)	152,089 (12.1%)	213,038 (16.9%)	1,260,460 (100.0%)

기본 어휘인 GSL의 경우 빈도 기준 1~1,000위에 해당하는 GSL 1,000의 비율이 66.1%, 1,001~2,000위에 해당하는 GSL 2,000의 비율이 4.9%로 합계 71.0%였다. 이러한 결과는 농업 분야의 학술 논문의 기본 어휘 비율을 67.5%로 보고한 Martínez et al.(2009)의 결과보다 약간 높으며, 다양한 학문 분야의 학술 텍스트의 기본 어휘 비율을 71.0%로 보고한 Coxhead(2000)의 결과와 유사한 결과이다. 한편 학술 어휘의 비율은 12.1%로 10%대를 기록했던 기존 선행 연구(Chen and Ge, 2007; Coxhead, 2000; Hyland and Tse, 2007; Li and Qian, 2010)에 비해 1~2% 가량 높은 비율을 보였다.

Coxhead와 Nation(2001)은 GSL과 AWL은 대체적으로 학술 텍스트의 90% 가량을 차지하며, 학술 텍스트를 원활히 읽기 위해서는 출현하는 어휘의 95% 가량을 알고 있어야 하기 때문에 5% 가량의 부족한 부분은 전문 어휘 학습을 통해 보완해야 한다고 제안하였다. 본 연구 결과에서도 컴퓨터 공학 학술 논문에서 GSL과 AWL의 출현 비율은 83.1%로 95% 수준에 크게 못 미치며, 이는 컴퓨터 공학 학술 텍스트에서 전문 어휘의 비율이 높을 가능성이 크고 동시에 컴퓨터 공학 분야의 EAP 학습자에게 전문 어휘의 학습이 더욱 중요함을 의미하는 결과라 할 수 있다.

### 4.2 전문 어휘의 추출

앞서 연구 방법에서 제시했던 것처럼 최소 7편 이상의 논문에서 총 36회 이상 등장한 어휘 중 기본 어휘인 GSL과 학술 어휘인 AWL에 속하지 않은 단어를 전문 어휘로 선정하였다. 그 결과 총 545개의 컴퓨터 공학 분야의 전문 어휘가 산출되었다. 전문 어휘로 선정된 어휘들을 빈도를 기준으로 약 50개씩 분류하여 총 11개의 목록으로 구분한 후 아래 <표 3>에 제시하였다. 즉 선정된 전문 어휘 중 목록 1의 어휘가 목록 2보다 높은 빈도를 기록한 것들이며, 같은 목록 내에서는 먼저 제시한 어휘가 뒤에 제시된 어휘보다 더 자주 등장한 것이다. 기존 선행연구 중 음성/음운론 분야의 전문 어휘를 추출한 이제영과 이해진(2019)의 연구에서는 372개, 영어교육학 분야의 전문 어휘를 분석한 이제영과 김정렬(2013)의 연구에서는 181개의 전문 어휘가 발견되었다는 점을 고려하면, 컴퓨터 공학 분야는 타 전공 분야에 비해 보다 다양한 전공 어휘가 활용되고 있었다.

표 3. 컴퓨터 공학 분야의 전문 어휘 목록

목록	어휘
1	algorithm, query, node, algorithms, cache, nodes, graph, execution, dataset, software, hardware, matrix, vector, apps, optimal, server, trajectory, clustering, storage, optimization, client, latency, lemma, vertex, cluster, theorem, architecture, candidate, linear, overhead, vertices, workload, score, database, app, runtime, baseline, bandwidth, queries, update, datasets, static, buffer, privacy, web, packet, budget, subset, scalability (49개)
2	iteration, benchmark, traffic, recall, configuration, online, processor, metrics, propagation, mapping, threshold, clients, column, loop, vectors, kernel, correlation, benchmarks, mobility, clusters, destination, binary, polynomial, chunks, novel, classifier, sensor, executed, sensors, graphs, servers, spatial, scores, iterations, tag, sparse, filter, configurations, heuristic, transaction, usage, profiles, updates, dependency, mobile, convergence, metric, reference, temporal, trajectories (50개)
3	settings, arithmetic, capture, solver, segments, interface, architectures, scan, semantics, locality, maximal, disk, verification, processors, encoding, dependencies, partitioning, semantic, parsing, segment, provisioning, audio, implementations, height, lexical, buffers, outlier, partition, linux, compression, circuit, additionally, syntactic, tags, efficiently, caches, array, annotations, cell, loops, outperforms, verify, matrices, cells, smart, bayesian, columns, scaling, filters, workloads (50개)
4	consecutive, execute, digital, auxiliary, chunk, updated, platform, parallelism, topology, integer, timing, slot, computes, feedback, fraction, generator, port, twitter, hybrid, switching, prefetching, chip, metadata, solvers, transactions, malicious, zone, synthetic, adversary, nonlinear, queue, engagement, packets, remote, optimized, default, video, correlations, phrase, compiler, simulations, decomposition, intel, references, gradient, optimize, ports, feasible, internet, placement (50개)
5	exponential, stochastic, bytes, chips, predictor, markov, factorization, recursive, objectives, simultaneously, suite, scheduler, slots, overview, coefficient, embedded, heuristics, prototype, synthesis, facebook, depicted, candidates, regression, synchronization, dependence, collaborative, checkpoint, multicore, diagonal, neural, smartphone, stack, proposition, annotation, google, sparsity, invocation, coefficients, automata, quantify, executing, classifiers, prefix, rewrite, smartphones, granularity, technologies, pipeline, executions (49개)
6	optimizing, iterative, differential, convex, intuitively, tuples, grid, exit, module, neighboring, robust, encode, partitions, slice, phrases, interconnected, wireless, java, plot, leverage, scalable, cumulative, inverse, footprint, thermal, tradeoff, outliers, sensing, weighted, mapped, session, thresholds, symmetry, counter, subjective, oblivious, profile, concrete, gaussian, caching, hash, aforementioned, filtering, annotated, workflow, deployment (48개)
7	magnitude, platforms, diagram, oracle, scoring, interleaving, captures, corollary, backup, notation, density, tuple, converge, subsets, rendering, outperform, overheads, corpus, van, sessions, tile, generalized, deployed, linguistic, parser, invocations, partitioned, competitive, databases, latent, proximity, primitives, cognitive, penalty, pairwise, genetic, positives, captured, functionality, kernels, desktop, encoded, arrays, schedulers, symmetric, vertical, baselines, byte (48개)
8	generic, spectrum, lemmas, vocabulary, collaboration, replicas, superior, heterogeneous, incoming, robustness, profiling, versus, reuse, latencies, predictive, predefined, frequencies, augmented, optimum, rankings, computationally, axis, plots, identifier, boolean, simultaneous, entropy, mac, accelerators, depicts, intuitive, switch, cardinality, embedding, destinations, correlated, updating, mathematical, infeasible, blackbox, duplication, dedicated, executes, interfaces, preprocessing, probabilistic, pseudocode, median, proxy, subsystem (50개)
9	verified, accelerate, converges, circuits, taxonomy, tablet, affinity, predicate, deterministic, histogram, mappings, intuition, offline, retrieval, spanning, homogeneous, synthesized, asynchronous, connectivity, atomic, iteratively, modules, optimizations, degradation, geometric, acceleration, informative, poison, facial, generalization, replication, simulator, programmer, dirichlet, negligible, configured, indices, layout, lookup, stationary, nonnegative, synchronized, collision, garbage, linearly, algorithmic, amazon, synchronous, concurrency, nominal, calibration (51개)
10	trivial, interestingly, experimentally, exponentially, traversal, rounding, pixels, sophisticated, redundant, clustered, verifying, encryption, compressed, legacy, centroid, huge, invalid, plausible, switches, multivariate, primitive, bottleneck, subsection, horizontal, statically, algebraic, capturing, feasibility, merge, descent, bayes, lab, slices, tolerance, retrieved, runtimes, segmentation, bin, sockets, conjunction, orthogonal, graphical, identifiers, programmers, ratings, predicates (46개)
11	encodes, initialization, parallelization, logarithmic, quantified, multithreaded, standardized, undirected, anomaly, xilinx, configurable, filtered, redundancy, architectural, counters, relational, disks, durations, suites, recursively, inspired, incurs, leveraging, dense, disjoint, alignment, pruning, parse, voltage, singular, conservative, instantiated, developer, naive, null, deploy, reused, engage, grammatical, assertion, parametric, referenced, contention, subsequence, discarded, leverages, suboptimal, artifacts, replicated, scalar, keywords, canonical, pixel, sketch (54개)



선정된 컴퓨터 공학 분야의 전문 어휘를 살펴보면 다음과 같은 특징이 있다. 첫째, 당연한 결과이지만 *algorithm, query, cache, node, software, hardware* 등 컴퓨터 공학과 관련성이 매우 높은 어휘들이 다수 포함되어 있었다. 둘째, *verify, objective, synthesis, effectively, reference* 등 컴퓨터 공학 분야를 포함한 다양한 전공 분야의 학술 텍스트에서 자주 등장할 것으로 예상되는 범학문적 학술 어휘들이 있었다. 셋째, *correlation, outlier, coefficient, regression, multivariate* 등 통계 관련 어휘들이 발견되었다. 넷째, *arithmetic, integer, fraction, factorization, algebraic* 등 수학 또는 연산 관련 어휘가 포함되어 있었다. 다섯째, *lemma, semantic, syntactic, corpus, vocabulary* 등 언어 또는 언어학 관련 용어도 전문 어휘 목록에서 발견할 수 있었다. 마지막으로 Intel, Google, Oracle, Xilinx와 같은 업체명, Linux, Facebook, Twitter와 같은 소프트웨어 및 서비스와 관련된 고유 명사로 등장하였다.

## 5. 결론

본 연구는 컴퓨터공학 전문어휘 목록을 제공하고자 컴퓨터 과학 분야의 국제 저명 학술지를 발간하는 기관인 ACM 에서 출간되는 27개의 학술지에서 게재된 총 135편의 학술논문을 대상으로 126만 단어 수준의 특수목적 코퍼스를 구축한 후, AntWordProfiler 1.4.1를 사용하여 전문어휘를 추출하였다. 컴퓨터 공학 특수목적 코퍼스의 전반적인 어휘 분포 양상을 분석한 결과 일반어휘인 GSL의 비율은 71.0% (GSL 1,000: 66.1%, GSL 2,000: 4.9%), 학술어휘인 AWL의 어휘분포 비중은 12.1%로 나타나 GSL과 AWL의 어휘를 모두 알고 있다면 약 83.1%의 어휘를 알고 있음을 의미한다. 이는 사전의 도움 없이 영어 읽기를 원활히 할 수 있는 수준인 95% 에 크게 미치지 못하는 것으로 컴퓨터공학 전문 텍스트 이해를 위해서 전문 어휘의 필요성이 특히 강조된다는 것을 알 수 있다. 컴퓨터공학 특수코퍼스에 등장하는 GSL 및 AWL를 제외한 어휘들 중 최소 7편 이상의 논문에서 총 36회 이상 등장한 어휘를 기준으로 총 545개의 컴퓨터공학 전문어휘가 산출되었다. 여타 학문 분야와 비교해볼 때 컴퓨터공학에 사용된 전문어휘의 수가 상대적으로 많다는 것을 확인할 수가 있는데 음성음운학(이제영·이혜진, 2019)에서는 372개, 영어교육학 분야의 전문 어휘(이제영·김정렬, 2013)는 181개, 유아교육학(이제영 등, 2017) 전문어휘는 224개, 공학분야(기계, 화학, 전기, 컴퓨터, 환경, 전자, 산업, 재료)의 공통 전문 어휘(하명호, 2017)는 357개로 컴퓨터 공학에 사용되는 전문어휘의 수가 높다는 것을 알 수 있다.

이러한 결과를 바탕으로 다음과 같은 교육적 활용 방안과 후속 연구를 위한 제언을 제안하고자 한다.

첫째, 컴퓨터공학에 등장하는 전문어휘들을 일상 생활과 연계지어 가르치는 것이다. 특정학문 분야에서 자주 등장하는 전문어휘는 해당 분야의 내용 및 개념과 밀접히 관련된 어휘들로 구성되어 있다. 전문어휘는 특정 학문에만 사용되는 영역이 아닌 일상생활에서 사용되는 어휘와 반드시 의미 차이가 있을 필요는 없고, 유일한 형식을 가질 필요도 없다. 전문 어휘가 지니는 본질적 특성은 의미적 측면에서 해당 학문 분야의 내용과 밀접한 관련이 있을 뿐이다. 예를 들어 본 연구에서 추출된 *apps, online, Google, offline* 등은 컴퓨터 공학에서는 전문용어로 분류되지만 일상 생활에서도 자주 사용된다. 이와 같은 의미적 연계성을 바탕으로 고빈도 어휘로 사용되는 어휘들과 컴퓨터 전문어휘의 유사성 및 차이점을 구별할 수 있도록 해야 한다. 예를 들면 일상생활에서는 교통의 뜻을 지닌 단어 *traffic*은 컴퓨터 전문분야에서는 서버에서 송수신되는 데이터의 전송량을 의미한다. 일상생활에서 사용되는 *traffic jam, heavy traffic* 등과 전문적으로 사용되는 *computer traffic, network traffic control* 등을 비교하여 관련 유사성을 파악하도록 하며 개념 및 지식 습득 차원에서 언어적 환원을 장려한다.

둘째, 자주 등장하는 전문어휘부터 직접적(direct)이고 명시적(explicit)인 교수-학습을 단계적으로 시행한다. 암시적, 우연적 어휘학습을 통해 특정 단어를 익히기 위해서는 최소 10회에서 12회 정도로 텍스트에 노출되어야 하는데 컴퓨터 용어와 같이 추상적이고 이질적인 용어는 해당 어휘의 배경지식이 없다면 직관적으로 이해하기가 어렵다. 전문어휘는 해당 학문분야의 핵심 어휘이자 동시에 내용어 역할을 하므로 문맥으로부터 뜻을 유추하기란 쉽지 않기 때문이다(이제영·김정렬, 2013). 전문어휘를 안다는 기본적인 전제는 그것이 결합되어 있는 지식의 총체를 관련지어 아는 것까지 포함한다는 것을 의미하는데 이는 곧 그 본질적인 면에서 학습의 중요성과 밀접히 관련되어 있다. 따라서 전문어휘의 개념적 측면을 명시적이고 직접적으로 지도-학습한 후 콘코던스 분석기능을 이용하여 해당 어휘가 쓰인 문장을 통해 목표 어휘에 익숙해지도록 장려한다면 전문어휘를 더 빠르게 습득할 수 있을 것이다.

셋째, 본 연구는 컴퓨터 공학 분야의 학술지 논문을 코퍼스로 구축하여 관련 전문 어휘를 분석하였다. 차후 연구에서는 학술지 논문 외의 자료들, 특히 대학 또는 대학원 수준에서 사용되고 있는 교재를 기반으로 전문 어휘 분석이 이루어질 필요가 있다. 또한 컴퓨터 공학 이외의 전문 어휘 분석이 이루어지지 않은 다양한 학문 분야에 대한 전문 어휘 분석이 진행될 필요가 있다.

본 연구는 시대적 변화 흐름을 반영하여 컴퓨터공학 분야를 위한 전문어휘 지식체계를 도출하고 목록으로 구조화하였다는 데 그 의의가 있다. 컴퓨터공학은 급격하게 기술이 발전되고 있는 분야로 사전에 실려 있어도 *floppy disk*와 같이 일부 어휘는 사어(dead language)가 되다시피 급격하게 변화하고 있다. 본 연구는 컴퓨터공학에서의 최신 쟁점 사항이나 국제적 기술 동향 등을 반영한 컴퓨터 분야의 국제저명 학술논문을 대상으로 전문어휘를 추출하였고 이를 토대로 컴퓨터 공학 영어어휘 교육에 필요한 토대를 구축하였다. 컴퓨터 공학과 같이 특히 학문의 개념 부하량과 용어의 특수성이 높은 학문에서는 어휘학습이 제대로 뒷받침되어야만 원활한 학습이 이뤄질 수 있는데, 본 연구에서 도출한 전문어휘 목록이 학습 효율을 극대화하는 방법을 모색하는 데 유용하게 사용되기를 희망한다.

## 참고문헌

- 설문원. 2003. 기록관리 용어 사용 실태 분석과 표준화 방안 연구. 『한국문헌정보학회지』 37.4, 241-268.
- 이제영·김정렬. 2013. 학술 전문 코퍼스를 활용한 영어교육 전문어휘 추출. 『Foreign Languages Education』 20.1, 99-122.
- 이제영·안종기·이지은. 2017. EAP 전문 코퍼스를 활용한 유아교육 전문어휘 추출. 『한국콘텐츠학회논문지』 17.1, 475-484.
- 이제영·이혜진. 2019. 음성학/음운론 분야 학술논문에서의 학술어휘 및 전문어휘의 분포: 코퍼스 기반 연구. 『언어과학』 26.1, 95-113.
- 장세은·변현정. 2011. 코퍼스를 활용한 해사영어 어휘 분석. 『새한영어영문학』 53.4, 247-268.
- 하명호. 2017. 공학영어 어휘 목록 개발. 『언어과학』 24.3, 167-185.
- 하명호. 2018. 영어교육 어휘 목록 개발: 국내외 학술지 코퍼스 기반 연구. 『영어영문학연구』 60.4, 377-398.
- 황성은. 2018. 대학 수학 목적 학습자를 위한 지식 기반 한국어 교육 연구-이른바 ‘범용 학술어휘’의 선정과 교육을 중심으로. 『국제한국어교육학회 춘계학술발표논문집』, 348-363.
- Anthony, L. 2014. *AntWordProfiler* (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/software>.

- Chen, Q. and G. Ge. 2007. A Corpus-based Lexical Study on Frequency and Distribution of Coxhead's AWL Word Families in Medical Research Articles (RAs). *English for Specific Purposes* 26.4, 502-514.
- Chung, T. M. and I. S. P. Nation. 2004. Identifying Technical Vocabulary. *System* 32.2, 251-263.
- Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34.2, 213-238.
- Coxhead, A. and I. S. P. Nation. 2001. The Specialized Vocabulary of English for Academic Purposes. In J. Flowerdew and M. Peacock (eds.), *Research Perspectives on English for Academic Purposes*. Cambridge, UK: Cambridge University Press, 252-267.
- Coxhead, A. and M. Demecheleer. 2018. Investigating the Technical Vocabulary of Plumbing. *English for Specific Purposes* 51, 84-97.
- Hyland, K. and P. Tse. 2007. Is There an "Academic Vocabulary"? *TESOL Quarterly* 41.2, 235-253.
- Hirsh, D. and I. S. P. Nation. 1992. What Vocabulary Size Is Needed to Read Unsimplified Texts for Pleasure. *Reading in a Foreign Language* 8.2, 689-696.
- Li, Y. and D. D. Qian. 2010. Profiling the Academic Word List (AWL) in a Financial Corpus. *System* 38.3, 402-411.
- Martínez, I. A., S. C. Beck, and C. B. Panza. 2009. Academic Vocabulary in Agriculture Research Articles: A Corpus-based Study. *English for Specific Purposes* 28.3, 183-198.
- Masrai, A. and J. Milton. 2018. Measuring the Contribution of Academic and General Vocabulary Knowledge of Learners' Academic Achievement. *Journal of English for Academic Purposes* 31, 44-57.
- Vongpumivitch, V., J. Huang, and Y. Chang. 2009. Frequency Analysis of the Words in the Academic Word List (AWL) and Non-AWL Content Words in Applied Linguistics Research Papers. *English for Specific Purposes* 28.1, 33-41.
- West, M. 1953. *A General Service List of English Words*. London: Longman.

이제영(제1저자), 교수  
 전북 전주시 완산구 천잠로 303  
 전주대학교 영어교육과  
 E-mail: jylee@jj.ac.kr

이혜진(교신저자), 교수  
 전북 익산시 익산대로 460  
 원광대학교 영어교육과  
 E-mail: lee.hyejin0419@gmail.com