



한국 성인 영어 학습자의 영작 장르별 어휘 빈도 프로파일과 어휘 사이즈*

강동호 (서울과학기술대학교)



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: September 07, 2021
Revised: September 25, 2021
Accepted: September 28, 2021

Dongho Kang
Professor, Dept. of English
Language and Literature, Seoul
Nat'l University of Science and
Technology
Tel: (02) 970-6250
dh14kang@seoultech.ac.kr

* This study was supported by the Research Program funded by the SeoulTech (Seoul National University of Science and Technology).

ABSTRACT

Kang, Dongho. 2021. Lexical frequency profile & vocabulary size in Korean adult EFL learners' writing genre. *Korean Journal of English Language and Linguistics* 21, 969-980.

The present study aims to investigate how different Korean adult EFL learners' productive vocabulary uses are among the writings of different genre. For the data analysis, AntWordProfiler was used to find Lexical Frequency Profile (GSL 1,000, GSL 2,000, AWL and others) and vocabulary size using British National Corpus (BNC)/Corpus of Contemporary American English (COCA) in the un-timed writing samples of definition, process, descriptive, and opinion genre. The findings showed that there were differences in LFP and vocabulary size per writing genre, especially, in GSL 2,000 word level. For example, GSL 2,000 words in descriptive writings were used more significantly than process and opinion ones, and those in definition writings more than opinion ones. In the vocabulary size of BNC/COCA per genre, 4,000, 5,000, 6,000, and 7,000 word levels were used in opinion, descriptive, process, and definition writing genres respectively in terms of about 98% of the word token used in writings. Pedagogical implications are discussed at the end.

KEYWORDS

lexical frequency profile, BNC/COCA, writing genre, AntWordProfiler, definition writing, process writing, descriptive writing, opinion writing

1. 서론

외국어 학습에서 어휘의 역할은 중요하며, 모든 의사소통 능력에서 어휘 능력은 필수적인 요소로 간주된다(Nation 2001). 학습자의 어휘 능력에 대한 평가는 수용적(receptive) 어휘 지식, 능동적(active) 어휘 지식, 어휘 빈도 프로파일(Lexical Frequency Profile, LFP)의 세 가지 관점에서 연구되었는데, 전자의 두 가지 평가 방식은 Vocabulary Levels Test와 Productive Levels Test와 같은 독립적인 평가 도구를 사용하고, 후자(LFP)는 실제 학습자의 영작에 나타난 어휘 사용을 통해서 평가하였다(Nation 2001).

어휘빈도 프로파일은 영작이나 발화의 의사소통 상황에서 사용된 학습자의 어휘를 평가한다는 점에서, 실제 어휘 사용 능력을 평가할 수 있다는 점에서 장점이 있다(Laufer and Nation 1995). Laufer와 Nation(1995)이 제안한 어휘 빈도 프로파일(LEP)은 학습자가 실제 영작에 사용된 어휘를 일반적 어휘 목록(General Service List, GSL)(West 1953)과 학술적 어휘 목록(Academic Word List, AWL)(Coxhead 2000)의 빈도를 분석하도록 고안되어있고, 이러한 어휘 사용 빈도는 다른 어휘 평가 도구와 상관관계가 있다는 것을 보여준다.

최근 코퍼스 연구와 코퍼스(corpus) 활용 어휘 분석 프로그램의 일반화로 실제 학습자의 어휘 사용 빈도를 측정하는 것이 용이하게 되었다. 예를 들면, RANGE(Nation and Heatley 2002)와 Anthony(2015)의 AntWordProfiler가 무료로 사용 가능한 프로그램이다. 이러한 프로그램은 코퍼스 자료를 분석하여 활용 빈도가 높은 어휘 목록을 발견하는 데 일조를 했다. 예를 들면, 어휘 빈도 프로파일에 사용된 2,000개의 빈도가 높은 일반적 어휘 목록(GSL)과 학문적 맥락에서 학생들에게 가장 필수적인 학술적 어휘 목록(AWL) 등은 학습자의 어휘 학습이나 교재 분석에 유용한 도구로 사용된다. 최근에는 영국영어와 미국영어의 광범위한 구어체 및 문어체의 코퍼스인, British National Corpus(BNC)와 Corpus of Contemporary American English(COCA)이 개발되어, 빈도에 따른 어휘 목록이 개발되었다(Nation 2006, Webb and Nation 2017).

최근에 실제 문맥에서 사용될 때 의사소통 능력이나 문맥에 적절한 어휘 사용 능력을 평가할 수 있다는 점에서 영작 등의 문어체에서 학습자가 사용하는 어휘를 분석하는 경향이 늘어나고 있다. 이러한 분석들은 주로 유창성(Higginbotham and Reid 2019, Ibrahim, Muhamad, Esa 2019), 과정 중심 영작(Muncie 2002), 작문 주제(윤현숙 2010, Kim and Ryoo 2011, Ryoo 2018)에 따른 어휘 빈도의 차이점에 초점을 둔 반면에, 대학교에서 학술적 영작의 장르에 따른 어휘 사용의 차이점에 관한 연구는 전무하다(Csomay and Prades 2018). Beglar와 Hunt(1999)는 대학에서 학문적으로 성공하기 위해서 영작에서 2,000개의 고빈도 어휘와 학문적 어휘(AWL)의 사용은 결정적인 역할을 한다고 주장한다. 이러한 점은 학문적 영작은 다양한 장르의 영작을 요구하고, 이에 적절한 어휘 사용이 수반된다는 것을 알 수 있다. 이에 본 연구는 영어 학습자의 학술적 영작문에 사용된 어휘 빈도와 장르에 따른 차이점을 분석하고자 한다.

2. 문헌 연구

Laufer와 Nation(1995)은 학습자의 영작에 나타난 어휘의 양(lexical richness)의 평가 도구로

서 어휘 빈도 프로파일(LEP)을 제안하였는데, 제 2언어 학습자의 영작에 나타난 어휘 빈도 프로파일의 점수는 다른 어휘 평가 도구의 점수와 비례한다는 것을 보여준다. 어휘 빈도 프로파일은 학습자의 영작에 나타난 일반적 어휘 목록(GSL)과 학문적 어휘 목록(AWL)의 비율 및 빈도를 조사하는데 구체적인 목적을 두고 있다. 이 연구는 동일한 학습자의 경우에 여러 장르의 영작에 사용된 어휘 양은 일정한 비율로 나타나고, 학습자의 유창성을 구분하는데 사용될 수 있다는 것을 보여준다.

Goodfellow, Lamy와 Jones(2002)는 어휘 빈도 프로파일(LEP)을 사용하여, 프랑스어 학습자의 작문 평가에 적용하였다. 개방대학(Open University) 프랑스어 과정에서 학습자의 텍스트를 분석하고, 이 결과를 강의 조교가 채점한 점수와 비교하였다. 그 결과 어휘 빈도 프로파일의 점수와 강의 조교가 부여한 점수와 유의미한 상관관계가 있었다. 이 연구자들은 어휘 빈도 프로파일의 결과가 학습자의 자아 평가의 피드백으로 사용할 것을 제안하였다.

Webb과 Nation(2017)도 학습자의 작문에 나타난 일반적 어휘(GSL)와 학술적 어휘(AWL)의 사용 비율을 조사하는 어휘 빈도 프로파일(LFP)을 통해서 학습자의 생산적 어휘 능력을 측정할 수 있다고 제안하였다. West(1953)는 영어 어휘에서 2,000개의 가장 일반적인 빈도가 높은 단어 목록(GSL)을 제시하였으면, 대부분의 텍스트에 나타나는 단어의 약 80%를 포함한다. 다른 한편으로, Coxhead(2000)는 West(1953)의 일반적 어휘 목록(GSL) 외에 3.5백만 어휘의 학문적 텍스트의 코퍼스 자료를 분석하여, 동일한 사이즈의 소설 코퍼스에서는 1.4% 정도만 차지하지만, 학술적 텍스트의 약 10%를 차지하는 어휘 목록(AWL)을 발견하였다. 이 목록은 대학에서 학문적으로 공부하는 학습자들에게 필수적인 어휘로서 여겨진다. Webb and Nation(2017)의 연구는 네 그룹의 어휘 목록(GSL 1,000, GSL 2,000, AWL, 목록 외)을 이용하여, 학습자의 작문에 사용된 어휘의 빈도를 분석하였다. 이러한 학습자의 어휘 빈도 자료(LFP)는 학습자의 어휘 능력을 신뢰할 수준으로 평가할 수 있고, 수준별 차이점도 가능하다는 것을 보여준다. 또한 어휘 빈도 프로파일(LFP)의 결과는 독립적인 어휘 평가 결과와 높은 상관관계를 보여주었다.

Higginbotham과 Reid(2019)는 Webb과 Nation(2017)의 어휘 빈도 프로파일(LFP)의 업데이트된 버전에 대한 타당성을 검증하기 위해서, 472명의 제2 외국어 학습자가 작성한 에세이 프로파일(LFP)을 분석하였다. 이전 연구와 달리 실제 환경에서 작성된 긴 에세이(2000 단어 길이)가 사용되었습니다. 연구의 결과는 학습자의 어휘 빈도 프로파일(LFP)이 작문 숙련도와 높지는 않지만 유의미한 상관관계를 보여주었다. 즉, 작문 점수가 낮은 학생들이 자주 사용되는 단어에 의존하고 상위 학습자들은 더 빈도가 낮은 어휘의 사용이 높다는 것을 보여준다. 그러나 어휘 빈도 프로파일(LFP)이 생산적 어휘 사용에 대한 평가로 신뢰할 수 있는 경향을 보이지만, 독립적인 작문 숙련도의 예측 변인으로 사용될 만큼 강력하지 않다는 것을 보여준다.

윤현숙(2010)은 한국의 대학생들을 대상으로 세 가지 다른 주제를 주고 과정 중심의 초안(first draft), 최종안(final draft), 시간 내 및 자유 영작을 실시한 후에 어휘 빈도 프로파일(LFP)을 통해 생산적 어휘 사용 사이즈를 조사하였다. 학습자의 어휘 빈도 프로파일(LFP)은 주제별로 차이점을 보이지만, 과정 중심 영작 수업에서 초안과 최종안에서는 어휘 사용 빈도에서 차이점이 없다는 것을 발견하였다.

Kim과 Ryoo(2011)는 독해 중심의 작문 수업에서 작문 주제(외모에 의한 사람 판단과 범인의 신상공개)와 읽기 및 쓰기 유창성에 따른 한국 대학생의 생산적 어휘 사용을 조사하였다. 학습자

의 어휘 사용은 네 가지 범주, 즉 GSL 1,000 단어, GSL 2,000 단어, 학문적 어휘(AWL), 나머지 어휘 군으로 나누어서 비교 분석하였는데, 독해력이나 영작 유창성은 어휘 사용에 차이점이 없었지만, 영작의 주제는 차이점을 주는 것으로 나타났다. 예를 들면, 외모에 의한 사람 판단 주제가 더 많은 1,000 단어와 내용어(content words) 사용을, 범인의 신상 공개 주제가 더 많은 2,000 단어 어휘와 범주 외 단어를 사용하게 유도하였으며, 상위 학습자일수록 범인의 신상 공개 주제에서 더 많은 기능어(function words)를 사용하는 것으로 나타났다.

Muncie(2002)는 일본 영어 학습자의 시간 내 영작과 과정 중심 초안과 최종 영작에 나타난 어휘 빈도 프로파일(LFP)을 분석하였다. 세 개의 영작에서 어휘 빈도 프로파일(LFP)의 유의미한 차이점은 없지만, 과정 중심 수업에서 초안보다 최종안에서 좀 더 섬세하고 상위 어휘의 비율이 높다는 점을 발견하였다. 그래서 과정 중심 교수법의 장점과 아울러 작문 전 단계에서 좀 더 명시적인 어휘 학습이 필요하다는 점을 지적하고 있다.

Ryoo(2018)의 최근의 연구는 영작 주제가 어휘의 다양성과 섬세함에 미치는 효과를 조사하였다. 어휘의 다양성과 사이즈는 어휘 종류와 토큰의 비율(type-token ratio)과 어휘 빈도 프로파일(LFP)로 각각 조사하였다. 윤현숙(2010)과 Kim and Ryoo(2011)의 결과와 유사하게, 작문의 주제가 어휘의 다양성과 사이즈에 영향을 주지만, 텍스트의 길이는 아무런 영향이 없는 것으로 나타났다. 즉, 영작의 주제가 어휘 선정이나 사용에서 중요한 역할을 한다는 것을 보여준다.

Ibrahim 외(2019)는 대학교 1학년생의 139개 에세이와 3학년의 에세이 140개 내용 분석을 통해 학습자의 생산적 어휘 사이즈를 조사하였다. 구체적으로, 이 연구는 GSL 1,000, GSL 2,000, 학문적 어휘(AWL), 목록에 없는 단어의 사용에서 두 집단 사이에 차이점을 Nation and Heatley(2002)가 개발한 RANGE 프로그램을 사용하여 조사하였다. 연구 결과에 따르면 두 그룹의 에세이는 두 그룹의 학습자에 의한 1,000, 2,000, 단어 수준 및 AWL 사용에서 통계적으로 유의미한 차이를 보여준다고 보도합니다. 그러나 본 연구의 결과를 자세히 보면, 1,000 단어 목록과 학술적 어휘(AWL)에서는 3학년이 더 사용빈도가 높고, 2,000 단어 목록에서는 오히려 1학년의 사용 빈도가 높아서 학습자의 수준이나 연령에 대한 척도로는 조금 어려움이 있어서 추가적인 논의가 필요하다.

Csomas와 Prades(2018)는 고등 교육 교실 환경에 있는 비원어민 학습자의 학문적 어휘 사용과 학생의 글쓰기 성적의 상관관계를 조사하고, 추가적으로 과제, 초안 및 수업 수준에 따른 학문적 단어 사용의 차이를 연구하였다. 결과를 보면, 일부 텍스트 유형에서 학문적 어휘 사용과 에세이 점수 사이에 유의미한 상관관계가 있으며, 초안과 관계없이 텍스트 유형 및 수업 수준에서 학문적 어휘가 사용되는 방식의 차이를 보여주었습니다. 그래서 수사적(rhetorical) 목적이 학습자의 작문에서 사용되는 학문적 어휘의 양에 중요한 영향을 미친다는 것을 보여주었습니다. 그래서 학생들은 과정 중심의 작문에서 일련의 초안과 수정본(revision)에서는 차이점이 없고, 수사적 목적에 따른 장르의 유형, 응답 논문(response paper), 비교 분석(comparative analysis), 탐색적 종합(exploratory synthesis), 논증 종합(argument synthesis), 수사적 분석(rhetorical analysis) 및 사설(editorial)에 따라서 어휘 사용의 사이즈가 달라진다는 것을 알 수 있다. 그러므로 과정 중심의 작문에서 초안 간의 차이에 따른 어휘 교육보다는 장르의 차이점에 따른 어휘 학습이 필요하다는 것을 보여준다.

3. 연구 방법

3.1 연구 주제

본 연구의 주제는 다음과 같다.

- (1) 장르별 영작문(정의, 과정, 묘사, 의견)에 사용된 어휘 빈도가 네 영역(GSL 1,000, GSL 2,000, AWL, 목록 외)에서 차이점이 있는가?
- (2) 장르별 영작문(정의, 과정, 묘사, 의견)에 사용된 학습자의 어휘 사이즈에서 차이점이 있는가?

3.2 연구 절차 및 도구

본 연구에 참여한 학습자는 서울 소재 평생교육대학의 일환으로 정규 과정에 재학 중인 16명의 성인 학습자로 2020년 2학기에 영작문의 기초 강좌를 수강하고 있다. 영어 유창도는 케임브리지 온라인 진단 평가(Cambridge-online test)에서 초급 13명(A1과 A2, 점수 6—17), 중급 1명(B2, 20점), 고급 2명(C1과 C2, 점수 21—24)으로 대부분이 초급 수준의 학습자이다. 학습자들은 수강 기간 동안, 정의(Definition), 묘사(Descriptive), 과정(Process), 의견(Opinion)에 해당하는 문단 단위의 영작을 제출하였다.

각 장르별 에세이는 AntWordProfiler(Anthony 2015)를 사용하여 연구 주제 1에 답하기 위해서 GSL 1,000, GSL 2,000, 학술적 어휘(AWL), 목록에 없는 단어의 네 개의 그룹으로 나누어 그 빈도를 분석하였다. 두 번째 연구 주제인 장르별 학습자의 어휘 사이즈는 BNC와 COCA의 빈도에 따른 10,000개의 어휘 목록을 사용하여 분석하였다(Nation 2006, Webb and Nation 2017).

연구주제 1의 분석하기 위해서 첫 번째 질문에 답하기 위해 SPSS 25 버전을 사용하여 장르별 차이점을 분석하기 위해서 다변량 일원분석(MANOVA)을 사용하였다. 어휘 빈도 프로파일(LFP)에 나타난 종속변인(GSL 1,000, GSL 2,000, AWL 등)의 차이점이 나타난 경우에는 개별 집단의 차이점을 사후분석(Post-hoc)을 통해서 검증하였다. 연구주제 2는 기술통계를 통해서 각 장르별 어휘 사이즈를 비교 분석하였다.

4. 결과

4.1. 영작문 장르(genre)별 어휘 빈도 프로파일

표 1은 정의에 해당하는 장르를 가지고 학습자마다 다양한 주제로 영작을 한 결과이다. 학습자들의 정의 영작의 어휘 사용(token)은 빈도수가 높은 GSL 1,000 어휘 목록에서는 77.34%, GSL 2,000에서는 8.10%, 학문적 어휘 목록에서는 5.01%를 사용하였다. 그래서 이 세 어휘 목록에서 90.45%의 어휘를 사용하였고, 목록 외의 어휘는 9.55%로 다소 높은 것으로 나타났다.

표 1. 정의(Definition) 영작의 어휘 빈도 프로파일

| 어휘 목록 (1,000) | Tokens N(%) | Types N | Families N | Cum(%) |
|---------------|----------------|------------|---------------|--------|
| GSL 1 | 1,126(77.34) | 364 | 288 | 77.34 |
| GSL 2 | 118(8.10) | 66 | 57 | 85.44 |
| AWL | 73(5.01) | 57 | 48 | 90.45 |
| 목록 외 | 139(9.55) | 81 | 81 | 100 |
| Total | 1,456(100) | 568 | 474 | 100 |

표 2의 어휘 빈도 프로파일 결과는 묘사(Descriptive)의 장르에 해당하는 영작에서 사용한 어휘(token)의 빈도를 보여준다. 어휘 빈도의 결과를 보면, GSL 1(81.28%), GSL 2(10.44%), 학문 어휘(1.97%), 목록 외(6.31%)를 보여준다.

표 2. 묘사(Descriptive) 영작의 어휘 빈도 프로파일

| 어휘 목록 (1,000) | Tokens N(%) | Types N | Families N | Cum(%) |
|---------------|----------------|------------|---------------|--------|
| GSL 1 | 828(81.28) | 286 | 234 | 81.28 |
| GSL 2 | 106(10.44) | 65 | 60 | 91.72 |
| AWL | 20(1.97) | 16 | 13 | 93.69 |
| 목록 외 | 64(6.31) | 48 | 48 | 100 |
| Total | 1,015 | 415 | 355 | 100 |

표 3의 결과는 과정(Process) 영작의 사용 어휘 빈도로서, 사용빈도 1,000 어휘군(80.55%), 2,000 어휘 군(7.66%), 학문적 어휘(3.44%), 목록 외(8.36%)를 보여준다.

표 3. 과정(Process) 영작의 어휘빈도 프로파일

| 어휘 목록 (1,000) | Tokens N(%) | Types N | Families N | Cum(%) |
|------------------|----------------|------------|---------------|--------|
| GSL 1 | 1,031(80.55) | 315 | 247 | 80.55 |
| GSL 2 | 98(7.66) | 74 | 68 | 88.21 |
| AWL | 44(3.44) | 29 | 27 | 91.65 |
| 목록 외 | 107(8.36) | 68 | 68 | 100 |
| Total | 1,280(100) | 486 | 410 | 100 |

마지막으로 의견(Opinion) 영작의 경우에는, 1,000(82.11%), 2,000(7.06%), 학문적 어휘(4.87%), 목록 외(5.96%)의 어휘 사용 빈도를 나타낸다.

표 4. 의견(Opinion) 영작의 어휘 빈도 프로파일

| 어휘 목록 (1,000) | Tokens N(%) | Types N | Families N | Cum(%) |
|---------------|----------------|------------|---------------|--------|
| GSL 1 | 2,850(82.11) | 614 | 426 | 82.11 |
| GSL 2 | 245(7.06) | 142 | 119 | 89.17 |
| AWL | 169(4.87) | 127 | 104 | 94.04 |
| 목록 외 | 207(5.96) | 127 | 127 | 100 |
| Total | 3,471(100) | 1010 | 776 | 100 |

위의 어휘 사용 빈도에 대한 결과를 보면, 장르별 차별성이 있다는 것을 알 수 있다. 이러한 차이점이 통계적으로 유의미한 지를 조사하기 위해서 표 5와 같이 다변량 일원분산분석(MANOVA) 분석을 실시하였다.

표 5. 다변량 일원분산분석

| 소스 | 종속변수 | 제3유형 | 제공합 | 자유도 | 평균 | 제공 | F | 유의 확률 |
|----|-------|-------|-----|-----|------|----|------|-------|
| 장르 | 1,000 | 18.31 | | 3 | 6.10 | | .21 | .889 |
| | 2,000 | 1.74 | | 3 | .58 | | 3.97 | .027* |
| | AWL | 1.22 | | 3 | .41 | | 2.64 | .085 |
| | 목록 외 | 1.74 | | 3 | .58 | | 2.80 | .074 |

Note. * $p < .05$

표 5의 결과를 보면, 장르별로 GSL 2,000 단어 목록에서 유의미한 차이점이 있다는 것을 보여준다. 추가적으로 종속 변수에서 개별 장르별 차이점을 조사하기 위해서 아래 표 6과 같이 사후 분석(Post Hoc Test)을 실시하였다.

표 6. 다중비교(2,000 어휘 목록)

| 종속변수 | 장르 | 장르 | 평균 차이 | 유의 확률 |
|----------------|----|----|-------|-------|
| 2,000 어휘 목록 | 정의 | 묘사 | -.07 | .78 |
| | | 과정 | .49 | .061 |
| | | 의견 | .61* | .023* |
| | 묘사 | 과정 | .56* | .035* |
| | | 의견 | .68* | .013* |
| | | 과정 | .12 | .626 |

Note. * $p < .05$

표 6에서 장르별 다중 비교를 한 결과, GSL 2,000 어휘 목록에서의 차이점은 정의, 묘사, 의견의 장르에서 나타나고 있다. 표 1-4의 결과처럼, 2,000 어휘 목록에서는 의견(7.06%), 과정(7.66%), 정의(8.10%), 묘사(10.44%)의 순서로 사용 비율을 보여준다. 그래서 통계상 유의미한 차이점으로, 정의(Definition) 영작은 의견(Opinion) 영작보다, 묘사(Descriptive) 영작은 과정(Process)이나 의견(Opinion) 영작보다 더 많은 2,000 어휘군을 사용한 것으로 나타났다.

4.2. 영작문 장르(genre)와 어휘 사이즈

표 7-10은 각 장르별 영작에 사용된 어휘 사이즈를 보여준다.

표 7. 정의(Definition) 영작의 어휘 사이즈

| 어휘 목록 | Tokens | Types | Families | Cum (%) |
|-------|--------------|-------|----------|---------|
| | N(%) | N | N | |
| 1 | 1,084(74.45) | 326 | 265 | 74.45 |
| 2 | 182(12.50) | 119 | 97 | 86.95 |
| 3 | 86(5.91) | 68 | 62 | 92.86 |
| 4 | 23(1.58) | 20 | 18 | 94.44 |
| 5 | 34(2.34) | 15 | 12 | 96.78 |
| 6 | 11(0.76) | 9 | 9 | 97.54 |
| 7 | 11(0.76) | 4 | 4 | 98.3 |
| 8 | 20(1.37) | 3 | 3 | 99.67 |
| 9 | 1(0.07) | 1 | 1 | 99.74 |
| 10 | 1(0.07) | 1 | 1 | 99.81 |
| 목록 외 | 3(0.21) | 2 | 2 | 100 |
| 합계 | 1,456(100) | 568 | 474 | 100 |

먼저 표 7에서는 98%의 기준 어휘 사이즈를 중심으로 학습자들은 대략 7,000 단어 수준까지의 어휘(98.3%)를 정의 영작에 사용한 것으로 나타난다. 이에 반해서, 표 8의 묘사 영작에서는 학습자들은 98.81% 수준에서 5,000 단어 수준까지의 어휘를 사용한 것으로 나타난다.

표 8. 묘사(Descriptive) 영작의 어휘 사이즈

| 어휘 목록 | Tokens | Types | Families | Cum (%) |
|-------|------------|-------|----------|---------|
| | N(%) | N | N | |
| 1 | 890(87.68) | 316 | 262 | 87.68 |
| 2 | 56(5.52) | 47 | 42 | 93.2 |
| 3 | 24(2.36) | 18 | 18 | 95.56 |
| 4 | 16(1.58) | 10 | 9 | 97.14 |
| 5 | 17(1.67) | 13 | 13 | 98.81 |
| 6 | 5(0.49) | 4 | 4 | 99.3 |
| 7 | 1(0.10) | 1 | 1 | 99.4 |
| 8 | 2(0.20) | 2 | 2 | 99.6 |
| 9 | 3(0.30) | 3 | 3 | 99.9 |
| 10 | 1(0.10) | 1 | 1 | 100 |
| 목록 외 | 0 | | | |
| 합계 | 1,015(100) | 415 | 355 | 100 |

표 9와 10에서는 과정과 의견의 영작에 사용된 어휘 사이즈를 보여주는데, 과정 영작에서는 6,000 단어 수준(98.83%)에서, 의견 영작에서는 4,000 단어 수준(98.21%)의 어휘를 사용하였다.

표 9. 과정(Process) 영작의 어휘 사이즈

| 어휘 목록 | Tokens | Types | Families | Cum (%) |
|-------|--------------|-------|----------|---------|
| | N(%) | N | N | |
| 1 | 1,026(80.16) | 322 | 257 | 80.16 |
| 2 | 131(10.23) | 78 | 67 | 90.39 |
| 3 | 54(4.22) | 34 | 33 | 94.61 |
| 4 | 24(1.88) | 22 | 20 | 96.49 |
| 5 | 19(1.48) | 12 | 9 | 97.97 |
| 6 | 11(0.86) | 6 | 6 | 98.83 |
| 7 | 5(0.39) | 4 | 4 | 99.22 |
| 8 | 3(0.23) | 3 | 3 | 99.45 |
| 9 | 2(0.16) | 2 | 2 | 99.61 |
| 10 | 4(0.31) | 2 | 2 | 99.92 |
| 목록 외 | 1(0.08) | 1 | 1 | 100 |
| 합계 | 1,280(100) | 486 | 404 | 100 |

종합하면, 각 장르별 어휘 사이즈를 보면, 기술적 수치만으로 보더라도 차이점이 있다는 것을 보여준다. 98% 수준에서 의견(4,000 단어), 묘사(5,000 단어), 과정(6,000 단어), 정의(7,000 단어)를 사용하였다. 장르마다 각각 서로 다른 어휘 사이즈를 사용한다는 것을 알 수 있다.

표 10. 의견(Opinion) 영작의 어휘 사이즈

| 어휘 목록 | Tokens | Types | Families | Cum (%) |
|-------|--------------|-------|----------|---------|
| | N(%) | N | N | |
| 1 | 2,792(80.44) | 586 | 423 | 80.44 |
| 2 | 401(11.55) | 235 | 185 | 91.99 |
| 3 | 177(5.10) | 116 | 96 | 97.09 |
| 4 | 39(1.12) | 28 | 26 | 98.21 |
| 5 | 29(0.84) | 21 | 19 | 99.05 |
| 6 | 6(0.17) | 6 | 6 | 99.22 |
| 7 | 13(0.37) | 7 | 7 | 99.59 |
| 8 | 4(0.12) | 4 | 4 | 99.71 |
| 9 | 1(0.03) | 1 | 1 | 99.74 |
| 10 | 2(0.06) | 2 | 2 | 99.8 |
| 목록 외 | 7(0.20) | 4 | 4 | 100 |
| 합계 | 3,471(100) | 1,010 | 773 | 100 |

5. 결론

본 연구의 결과를 보면, 먼저 영작문 장르에 따른 어휘 빈도 프로파일(LFP)은 차이점을 보여주고 있다. 기술통계에서, GSL 1,000 단어가 차지하는 비율은 의견(82.11%), 묘사(81.28%), 과정(80.55%), 정의(77.34%) 영작의 순서를 보여주며, GSL 2,000은 묘사(10.44%), 정의(8.10%), 과정(7.66%), 의견(7.06%)의 순서로, 특히 의견 영작에서 GSL 1,000이 더 많고, 정의 영작에서 GSL 2,000 단어의 비율이 높다는 것을 보여준다. 학술적 어휘(AWL)에서는 정의(5.01%), 의견(4.87%), 과정(3.44%), 묘사(1.97%)의 순서로 정의나 의견 영작에서 학술적 어휘의 비율이 높

다는 것을 알 수 있다. 통계적 유의미한 차이점은 GSL 2,000에서 나타나는데, 묘사 영작이 과정이나 의견 영작보다, 정의 영작은 의견 영작보다 더 유의미한 수준으로 차이점을 보여준다.

독해 지문에 비해서 영작의 경우에는 고빈도의 일반적 어휘(1,000 어휘 군)가 사용될 확률이 높지만, 2,000 어휘군의 사용에서 장르별 차이점은 교육적 시사하는 바가 있다. 즉 장르에 따른 어휘 사용의 정도와 선택이 다르다는 점은 교사가 학습자에게 각 장르에 적절한 어휘 학습의 필요성이 제기되며, 어휘 빈도 프로파일의 정보는 학습자에게 자아 평가의 용도로 활용하는 방법도 고려된다(Goodfellow 외 2002). 또한 학술적 어휘(AWL)의 장르별 영작문에서 사용 비율을 보면 전반적으로 낮다는 것을 알 수 있다. Coxhead(2000)는 학술적 어휘(AWL)가 학술 텍스트의 약 10%를 차지하지만, 일반 텍스트의 경우에는 약 1.4%의 낮은 비율을 차지한다고 발표하였다. 즉 학술적 어휘 목록은 대학 등에서 학문적으로 공부를 하는 경우에는 중요한 역할을 한다는 것을 알 수 있다. 그러나 본 연구에서는 전체 영작에서 5%미만의 사용 빈도를 보여준다는 점에서 학술적 어휘에 대한 명시적 학습이 요구된다는 것을 알 수 있다.

장르별 사용 어휘의 사이즈에서도, BNC와 COCA의 빈도에 따른 10,000개의 어휘 목록에서 차이점을 보여준다(Nation 2006, Webb and Nation 2017). 대략 98% 기준으로 의견(4,000), 묘사(5,000), 과정(6,000), 정의(7,000)의 순서로 사용된 어휘 사이즈가 구분된다는 것을 알 수 있다. 이러한 결과는 Csomay and Prades(2018)의 선행연구와 일치한다. 즉, 장르에 따라 사용 어휘의 사이즈가 다르다는 점을 보여주고 있다.

읽기에 관한 선행 연구는 적절한 독해력에는 텍스트의 95%에서 98%의 어휘 지식이 요구된다고 보여준다(Hu and Nation 2000, Laufer 1989). 학술적 텍스트인 경우에는 최적의 어휘 지식의 수치로 98%의 어휘 지식이 최적이라고 여겨진다(Schmitt, Jiang and Grabe 2011). 95%와 98%의 어휘 지식은 BNC와 COCA의 어휘 목록에서 고유명사를 포함하여, 각각 4,000-5,000 단어와 8,000개의 단어에 대한 지식이 요구된다(Laufer and Ravenhorst-Kalovski 2010). 위의 연구 결과는 영작문 장르별 4,000에서 7,000개의 어휘가 사용된다는 것을 보면, 영작문의 종류에 따라 학습자에게 요구되는 어휘의 수가 달라진다는 점을 보여준다.

이러한 결과는 장르별 어휘 학습이 달라져야 한다는 교육적 시사점이 제기된다. Nation(2001)에 의하면, 어휘 학습에서 목적에 따라 4가지 접근법을 제기한다. 이해 가능한 의미 중심의 입력(comprehensible meaning-focused input), 언어 초점 학습(language-focused learning), 의미 중심의 출력(meaning-focused output), 마지막으로 유창성 신장(fluidity development)이 여기에 해당한다. 어휘 빈도 프로파일(LFP) 분석은 바로 의미 중심의 출력에 해당된다고 볼 수 있다. 그러나 장르별 어휘 사이즈나 사용 빈도가 다르다는 점을 고려하면, 빈도가 낮고 난이도가 높은 장르의 경우에는 언어 형태 중심의 명시적 어휘 학습을 영작문 전후에 제공할 필요성이 있다.

참고 문헌

- 윤현숙(Yoon, H.). 2010. 한국 학생들의 영어 작문에 나타난 생산적 어휘에 관한 분석(An analysis of productive vocabulary in Korean learners' L2 writing). 《영어학》(*Korean Journal of English Language and Linguistics*) 10(1), 85-104.

- Anthony, L. 2015. [Computer Program]. <http://www.laurenceanthony.net/software/antwordprofiler/>
- Beglar, D. and A. Hunt. 1999. Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing* 16(2), 131–162.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2), 213–238.
- Csomas, E. and A. Prades. 2018. Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes* 33, 100–118.
- Goodfellow, R., Lamy, M. and G. Jones. 2002. Assessing learners' writing using lexical frequency. *ReCALL* 14(1), 133–145.
- Higginbotham, G. and J. Reid. 2019. The lexical sophistication of second language learners' academic essays. *Journal of English for Academic Purposes* 37, 127–140.
- Hu, M. and P. Nation. 2000. Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13(1), 403–430.
- Ibrahim, E. Muhamad, J. and Z. Esa. 2019. A comparison of lexical richness in L2 written productions. *International Journal of Emerging Technologies in Learning* 14(20), 174–181.
- Kim, S–Y. and Y. Ryoo. 2011. Korean EFL learners' vocabulary use in reading-based writing according to topic and learner proficiency. *English Teaching* 66(1), 91–109.
- Laufer, B. 1989. What percentage of text–lexis is essential for comprehension? In C. Lauren, and M. Nordman eds., *Special Language: From Humans Thinking to Thinking Machines*, 316–323. Bristol, UK: Multilingual Matters.
- Laufer, B. and P. Nation. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16(3), 307–22.
- Laufer, B. and G. Ravenhorst–Kalovski. 2010. Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language* 22(1), 15–30.
- Muncie, J. 2002. Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System* 30(2), 225–235.
- Nation, P. 2001. *Learning Vocabulary in Another Language*. New York: Cambridge University Press.
- Nation, P. 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review* 63(1), 59–82.
- Nation, P. and A. Heatley. 2002. *RANGE: A Program for the Analysis of Vocabulary in Texts*. <http://www.victoria.ac.nz/lals/about/staff/paulnation>
- Ryoo, Y. 2018. Comparing lexical diversity and lexical sophistication in Korean EFL writing: Topic and text length. *Multimedia-Assisted Language Learning* 21(3), 63–87.
- Schmitt, N. Jiang, X. and W. Grabe. 2011. The percentage of words known in a text and reading comprehension. *The Modern Language Journal* 95(1), 26–43.

Webb, S. and P. Nation. 2017. *How Vocabulary Is Learned*. Oxford: Oxford University Press.
West. M. 1953. *A General Service List of English Words*. London: Longman, Green & Co.

예시 언어(Examples in): 한국어(Korean)
적용 가능 언어(Applicable Languages): 영어(English)
적용 가능 수준(Applicable Levels): 대학(College)