

영미 아동 모험 소설에 관한 코퍼스 분석 연구: 『보물섬』을 중심으로

A Corpus Analysis of British-American Children's Adventure Novels: Treasure Island

최은샘*, 정채관**

인천대학교 코퍼스연구소*, 인천대학교 영어영문학과**

Eunsaem Choi(eschoi@icr.or.kr)*, Chae Kwan Jung(ckjung@inu.ac.kr)**

요약

본 연구에서는 대표적인 영미 아동 모험 소설 『보물섬』의 언어적 특징을 파악하기 위해 『보물섬』을 코퍼스화 하여 어휘, 리마, 키워드, n-그램을 분석하였다. 이 연구를 통해 고빈도 어휘가 텍스트의 핵심어라는 일반적인 주장과 달리 『보물섬』의 고빈도 어휘는 『보물섬』과 직접 관련이 없는 기능어, 고유명사 등이 최상위층에 포진하고 있다는 것을 발견하였고, 통계적인 방법으로 추출한 『보물섬』 키워드 역시 『보물섬』의 내용을 가늠하기에 충분하지 않음을 발견하였다. 따라서 1차 정량적인 키워드 분석 후 진행된 2차 정성적인 키워드 분석을 통해 추출한 30개의 핵심 키워드를 통해 『보물섬』 내용을 신속하고 구체적으로 파악하는 단초를 마련하였고, 이를 바탕으로 그동안 직관적으로만 회자 되던 『보물섬』에 나타난 남성성을 계량적으로 분석할 수 있었다. 또한, n-그램 분석을 통해 『보물섬』의 작가가 다른 작가에 비해 선호하고 자주 사용하는 연속어휘구를 발견하였고, 이를 토대로 문학 작품의 계량적 연구가 가능한 코퍼스 문체론 연구의 가능성을 탐색하였다. 본 연구를 통해 밝혀낸 연구결과가 영미 아동문학 콘텐츠의 확산과 코퍼스 문체론 연구에 도움이 되기를 희망한다.

■ 중심어 : | 문학 텍스트 | 코퍼스 문체론 | 어휘 분석 | 키워드 분석 | N-gram 분석 |

Abstract

In this study, we analyzed the vocabulary, lemmas, keywords, and n-grams in 『Treasure Island』 to identify certain linguistic features of this British-American children's adventure novel. The current study found that, contrary to the popular claim that frequently-used words are important and essential to a story, the set of frequently-used words in 『Treasure Island』 were mostly function words and proper nouns that were not directly related to the plot found in 『Treasure Island』. We also ascertained that a list of keywords using a statistical method making use of a corpus program was not good enough to surmise the story of 『Treasure Island』. However, we managed to extract 30 keywords through the first quantitative keyword analysis and then a second qualitative keyword analysis. We also carried out a series of n-gram analyses and were able to discover lexical bundles that were preferred and frequently used by the author of 『Treasure Island』. We hope that the results of this study will help spread this knowledge among British-American children's literature as well as to further put forward corpus stylistic theory.

■ keyword : | Literary Text | Corpus Stylistics | Lexical Analysis | Keyword Analysis | N-gram Analysis |

* 이 논문은 인천대학교 2019년도 융합과학기술원 신규융합연구단 지원에 의하여 연구되었음.

접수일자 : 2020년 09월 17일

심사완료일 : 2020년 10월 26일

수정일자 : 2020년 10월 26일

교신저자 : 정채관, e-mail : ckjung@inu.ac.kr

I. 서론

영미 문학 연구 분야에서 모험 소설은 통상 판타지 소설에 포함되어 광범위하게 분석되었다. 그 이유는 모험 소설에 판타지적인 설정이 자주 들어가 있기 때문이다. 예컨대, 널리 알려진 모험 소설 『피터팬』은 네버랜드라는 판타지 세계에서 펼쳐지는 이야기이다. 모험 소설은 현실과 흡사하지만, 판타지적 요소를 가미하여 현실에서 이루 수 없는 욕망을 표출하는 설정을 자주 사용한다. 독자들은 가상현실에서 현실의 제한된 틀을 벗어나 모험을 통해 상상력을 넓히고 모험심을 키우는 것이다. 상황이 이렇다 보니 모험 소설은 다른 소설 장르와 달리 주인공이 아동인 경우가 많고, 독자 역시 아동이 많다. 아동 모험 소설이 현실과 허구를 명확히 구분하지 못하는 시기인 어린 독자들에게 끼치는 영향이 적지 않게 되는 구조이다.

감동적이고 탄탄한 이야기 구조를 가진 다수의 영미 아동 모험 소설은 우리나라에서 교과서에 수록되기도 하고, 출판사와 도서관 추천 도서로 선정되는 등 영향력이 적지 않다[1]. 예를 들어 『보물섬』은 도전과 모험 이야기라는 주제로 초등학교 3학년 이상을 대상으로 세계 명작문학의 하나로 선정되어 있다. 하지만 이러한 영미 아동 모험 소설이 우리의 문화에서 허용되는 적절한 내용과 교육적 함의를 포함하는지에 관한 논의는 여전히 부족한 상황이다. 서구의 문화와 우리의 문화가 다르다는 점을 고려하면, 특히 어린 아동에 적지 않은 영향을 끼치는 영미 아동문학에 관한 정서와 문화적 측면을 더욱 능동적으로 연구·분석할 필요가 있는 것이다.

본 연구에서는 문학 작품을 언어 빅데이터로 만들어 통계적인 방법에 따라 도출된 정량적 결과를 정성적으로 추가 분석하는 코퍼스 접근방법을 활용하여 우리나라 많은 아동에게 읽히는 『보물섬』의 어휘, 리마, 키워드, N-그램을 토대로 작품의 특성과 작가의 스타일을 색다른 시각에서 살펴보고자 한다.

II. 이론적 배경

1. 영미 아동 모험 소설

아동이 주 독자층인 영미 아동 모험 소설은 문학의 한 장르로서 모험을 소재로 다룬 소설이다. 주인공의 생각과 행동을 중심으로 이야기가 전개되며 대중문학과 아동문학 분야에서 소개됐다. 모험 소설은 시대적 배경을 반영하는 경우가 많은데, 이러한 시대적 배경 역시 모험 소설의 성격과 발전에 영향을 끼쳐왔다[2]. 모험 소설은 특히 19세기를 기점으로 성격이 달라지는데, 19세기 전반기 모험 소설은 시대적 상황을 반영하여 당시 종교적·도덕적 가르침을 반영하는 사례가 많았지만, 후반기로 갈수록 가족 단위나 개인의 모험을 중심으로 그간의 당연한 가치가 아닌 새로운 가치를 탐색하는 내용으로 변화되기 시작하였다[3]. 이러한 분위기는 『걸리버여행기』, 『로빈슨크루소』, 『피터팬』, 『톰소여의모험』, 『정글북』 등에서 볼 수 있듯이, 주인공이 우연한 기회에 낯설고 기묘한 곳으로 모험을 떠나게 되는 내용에서 나타난다[4].

예컨대 『로빈슨크루소』는 주인공이 바다에서 태풍을 만나 배가 난파당한 뒤, 낯선 곳에서 모험이 시작된다. 위험할지도 모르는 낯선 곳으로의 모험은 주인공의 성격을 변화시키기도 한다. 원래는 평범한 성격이었으나, 예상치 못한 일과 위험한 일을 겪으며 규율과 질서에 순응하던 평범한 성격이 점차 기존의 삶과 거리가 있는 다소 과격한 행동과 결정을 하게 된다[5]. 이러한 주인공의 변화는 주인공에게 없었던 용기를 요구하기도 하고, 원래 성격과 다른 무자비함과 교활함까지 요구하기도 한다. 문제는 영미 아동 모험 소설에 나타난 이러한 주인공의 행동이 아직 신체적, 인지적, 정신적으로 미숙한 우리나라 아동에게 끼치는 영향이다.

아이는 어른이 되기까지 수많은 내면적 갈등과 정신적 성장을 체험한다[6]. 특히 세계적으로 널리 알려진 명작 아동 모험 소설은 탄탄한 구성과 줄거리를 바탕으로 주인공의 이상과 소망의 표현이 흥미진진하고 구체적인 이미지로 나타나기 때문에 이러한 환상의 간접 경험은 아동의 정체성 형성과정에서 중요하게 작용한다[7]. 따라서 지나치게 용감하고, 폭력적이고, 자칫 편향된 남성 우월주의 등은 아동 독자의 주의가 필요하다.

2. 보물섬

『보물섬』은 원래 1881년부터 ‘선박요리사’라는 제목으로 영국 잡지 ‘젊은 사람들’에 연재되다가, 1883년 책으로 출간되었다. 『보물섬』을 쓴 로버트 스티븐슨은 영국 출신 소설가로 어릴 때부터 몸이 약해 유럽의 여러 곳에서 요양 생활을 하였다. 스티븐슨은 자신의 이러한 신체적·환경적 영향으로 모험을 동경하게 되었고, 그러한 상황에서 쓴 글이 『보물섬』이다. 『보물섬』은 영국 빅토리아 시대의 대표적인 모험 소설이다. 당시 식민지 개척과 해외 무역에 관심이 많았던 영국은 바다 건너 바깥 세계를 탐험하고자 하는 욕구가 강했다[8]. 이러한 생각을 하고 있던 영국인들에게 바다는 방대한 미개척 공간으로 도전과 희망을 자극하고, 다른 한편으로는 신비한 미지의 공간이며, 또 다른 한편으로는 두려움을 불러일으키는 곳이다. 바다 건너 보물을 찾기 위해 모험을 떠나는 『보물섬』에는 이러한 시대적 배경과 당시의 분위기를 고스란히 녹여놓았다.

『보물섬』의 주인공 짐 호킨스는 13살의 평범한 소년이었지만, 보물섬으로의 모험을 통해 점점 용감하고 성숙하게 성장한다. 하지만, 이 과정에서 주인공에게는 남성 중심 세계관, 도덕성, 윤리 등 내면적 파괴도 함께 나타난다. 모험 소설에서 나타나는 이상적 행동 특성인 정직, 용기, 양심, 모험심과 같은 사항은 아동 독자의 바람직한 행동 특성에 긍정적으로 작용한다[9]. 어린 독자들은 『보물섬』의 주인공과 자신을 동일시하고, 주인공의 모험에 함께 참여하기 때문이다[10]. 따라서 특히 아이의 정체성이 형성되어가고 있는 아동기 독자에게 모험을 성공적으로 이끄는 용감한 주인공의 모습은 아동 독자가 자신의 가치관 형성을 돕는 중요한 역할을 한다.

한편, 문학 작품에서 전개되는 이야기에서 현실과 판타지를 구체적으로 구분하기는 쉽지 않다. 문학이 상상과 허구의 산물이기도 하지만 실제 존재하는 사람, 장소, 사물 등이 동시에 활용되기 때문이다[11]. 특히 모험 소설은 작품에 따라 실제 존재하는 나라나 지명 등 현실 세계를 바탕으로 현실 세계에는 없는 비가시한 환상의 세계를 보여주는 환상을 보여주기도 한다[12]. 모험 소설이 본질적으로 경이로움, 신비, 동경으로 가득

찬 미지의 모험을 다룬 소설이므로 판타지적 요소를 배놓을 수 없다는 점에서 아동 독자가 현실과 다른 세계에서 통용될 수 있는 것에 관한 구분 지을 수 있는 교육적 내용도 함께 비평적으로 읽을 수 있는 시각이 필요한 이유다.

3. 코퍼스 기반 문학 텍스트 분석 연구

코퍼스의 사전적 의미는 어떤 것의 모음(a collection of thing)이다. 인천대학교 코퍼스연구소에서는 이 어떤 것(anything)을 데이터(data)로 간주하고, 코퍼스를 데이터(텍스트, 이미지, 음성, 영상 등)를 모아 목적에 맞게 분류한 후 정제된 것이라고 정의한다[13]. 코퍼스에 관한 정의는 학자마다 조금씩 견해가 다르지만, 전반적으로 인간의 언어(문어와 구어)를, 컴퓨터에서 분석할 수 있는 전자문서 형태인 것을 의미한다는 데는 이견이 없다.

연구자들은 영미 문학 작품을 분석할 때 전통적으로 연구자의 직관으로 작품을 분석해왔다. 반면, 코퍼스 접근방법은 객관적인 상태에서 어휘적, 문법적, 맥락적 영미 문학 작품을 컴퓨터로 분석하고 결과를 도출한 뒤, 연구자의 경험으로 해석한다[14]. 코퍼스 기반 문학 작품 분석은 특히 작품 수가 많거나 규모가 방대할수록 그 효과성과 효용성이 빛을 발한다. 연구자 1인이 문학 작품에 등장하는 모든 등장인물의 감정 정보를 모두 해석하는 것은 물리적으로나 정신적으로 매우 소모적인 일일 뿐 아니라 연구자의 주관적인 해석 위주이기 때문에 확정적인 형태로 정의를 내리는 것에 한계가 있기 때문이다[15]. 컴퓨터를 사용한 정량적 분석과 정성적 혼합 방식은 연구자 직관으로 보기 어려운 문학 텍스트의 언어학적 패턴을 신속하고 정교하게 분석할 수 있다[16]. 또한, 코퍼스 분석을 통해 추출한 빈도수가 높은 어휘에 주목한다. 작품의 내용이나 구조에서 중요한 의미가 있을 수 있기 때문이다[17].

하지만 출현 빈도가 높다고 해서 분석 대상 텍스트에서 늘 중요한 의미가 있다고 할 수는 없다. 예컨대 기능어에 해당하는 관사나 전치사 등은 어떤 텍스트를 분석해도 늘 최고빈도 그룹에 속해 있기 때문이다. 따라서 빈도수가 두드러지고, 동시에 텍스트와 밀접한 관련이

있는지 키워드 분석을 통해 진짜 키워드를 도출할 필요가 있는 것이다. 즉, 1차 정량적인 고빈도 어휘 추출이나 키워드 추출 이후, 2차 정성적인 유의미한 고빈도 어휘나 키워드 선별작업이 이어져야 코퍼스 접근방법이 효과가 있는 것이다.

III. 연구방법

1. 분석 대상 및 분석 도구

본 연구에서는 『보물섬』 텍스트를 무료로 제공하는 프로젝트 구텐베르크(<http://www.gutenberg.org>)에서 『보물섬』 텍스트를 수집하여, 보물섬코퍼스를 구축하였다[18]. 본 연구에서 사용한 주요 분석 도구는 WordSmith Tools 7(이하, '워드스미스')와 AntConc 3.5.8(이하, '안트콘크')이다. 보물섬코퍼스의 전반적인 기초 통계 정보를 측정하기 위해 워드스미스를 사용하였고, 고빈도 어휘, 키워드, n-그램 추출 및 분석을 위해서는 워드스미스를 사용하였다.

2. 분석 과정

본 연구에서는 『보물섬』의 언어적 특성 파악을 위해 다음 항목을 측정하고 분석하였다. 첫째, 『보물섬』의 기초 통계 정보를 분석하였다. 어휘 다양성 측면에서 텍스트에서 추출된 타입과 토큰을 기반으로 타입토큰비율(Type-Token Ratio, 이하 'TTR')을 분석하였고, 더욱 정교한 TTR 분석을 위해 표준화된 TTR(Standardised Type-Token Ratio, 이하 'STTR')을 분석하여 실제적인 어휘 다양성 정도를 측정하였다[19]. 둘째, 『보물섬』에 사용된 어휘의 변화형을 리마 처리하여 기본형 목록을 추출하였다. 리마는 예를 들어, 기본형 *be*를 기준으로 *is, be, are, was, am, were, being, been* 등을 변화된 형태로 보고, 이들을 모두 'be'로 간주하여 처리한다. 이렇게 되면, 『보물섬』에 사용된 어휘의 기본형 목록을 추출할 수 있고, 이를 통해 기본 어휘와 개수를 파악하여 작가의 어휘 사용 스타일을 이해하는 데 도움이 된다. 본 연구에서 사용

한 리마 리스트는 와세다대학 로런스 앤서니 교수가 안트콘크 홈페이지를 통해 제공하는 Someya Lemma List (no hyphens)이다[20].

셋째, 1억 단어 규모의 영국국립코퍼스(British National Corpus, 이하 'BNC')를 참고코퍼스로 하여 『보물섬』의 키워드를 추출하여 분석하였다. 키워드는 보물섬코퍼스 자체에 사용된 어휘를 단순히 빈도순으로만 보여주지 않고, BNC와 비교하여 통계학적으로 유의미할 만큼 그 빈도가 유난히 높거나 낮은 결과를 바탕으로 도출된다[21][22]. 키워드는 연구 대상 코퍼스의 내용을 신속하게 파악하는 데 유용하다. 마지막으로 본 연구에서는 영미 아동 모험 소설의 특성과 작가의 스타일을 분석하기 위해 『보물섬』의 n-그램(gram)을 분석하였다. N-그램은 고빈도 어휘나 키워드처럼 1개의 어휘만으로는 보기 힘든 부분을 파악하는 데 도움이 된다[23-25]. 만일 특정 소설이나 학술지 등에서 특정 n-그램이 반복적으로 자주 사용된다는 것을 리마 리스트의 고빈도 어휘와 달리 저자의 무의식중에 자주 사용하는 저자의 정형화된 표현 방식이거나, 저자가 의식적으로 두드러지게 강조하고 싶은 연속어휘군일 가능성이 크다. 전자일 경우는 저자의 스타일을 연구하는 데 도움이 되고, 후자의 경우는 작품을 이해하는 데 도움이 된다. 이러한 접근은 코퍼스 문체론 분야에 적용되어 영미 문학 작품의 수많은 작가의 스타일을 분석하는 데 유용하게 활용되고 있다[26]. 작가가 선호하는 어휘를 바탕으로 작가를 구별할 수도 있지만, 연속어휘구를 바탕으로 작가를 추측할 수도 있다.

IV. 결과 및 논의

1. 보물섬코퍼스 기초 통계 분석

본 연구에서는 『보물섬』 텍스트를 무료로 제공하는 프로젝트 구텐베르크에서 텍스트를 수집하여 코퍼스와하였고, 워드스미스를 사용하여 기초 통계정보를 측정하였다. [표 1]은 보물섬코퍼스의 기초 통계정보를 정리한 것이다.

표 1. 보물섬코퍼스 기초 통계정보

분류	보물섬코퍼스
텍스트	1
토큰	68,717
타입	6,019
TTR	8.76
STTR	81.45

[표 1]에서 볼 수 있듯이, 보물섬코퍼스는 총 68,717 토큰으로 이루어져 있으며, 타입은 6,019개이다. 따라서 TTR은 8.76이며, 이를 표준화시킨 STTR은 81.45이다. 단순히 보물섬코퍼스의 숫자만을 나열하면, 보물섬코퍼스의 상황이 어떤지 파악하기 어려우므로 같은 방법으로 다른 영미 아동 모험 소설 『걸리버여행기』, 『로빈슨크루소』, 『정글북』, 『톰소여의모험』, 『피터팬』의 텍스트를 수집하여 영미아동모험소설코퍼스(Corpus of British-American Children's Adventure Novel, 이하 'CBACAN')를 구축하였고, CBACAN의 기초 통계 정보를 측정한 결과는 [표 2]와 같다.

표 2. CBACAN 기초 통계정보

분류	걸리버 여행기 코퍼스	로빈슨 크루소 코퍼스	정글북 코퍼스	톰소여의 모험 코퍼스	피터팬 코퍼스	합
텍스트	1	1	1	1	1	5
토큰	104,932	121,733	52,344	72,206	47,707	398,922
타입	8,149	6,074	4,577	7,481	4,951	16,323
TTR	7.77	4.99	8.74	10.36	10.38	4.09
STTR	82.59	78.51	81.01	80.58	81.81	80.59

CBACAN과 보물섬코퍼스를 비교해보면, 우선 보물섬코퍼스의 토큰 수가 68,717개이고, CBACAN의 토큰 평균이 79,784개이다. 『보물섬』이 『정글북』이나 『피터팬』보다는 토큰 수가 많지만, 그 밖의 영미 아동 모험 소설 『걸리버여행기』, 『로빈슨크루소』, 『톰소여의모험』보다는 토큰 수가 적다는 것을 알 수 있다. 즉, 이러한 결과는 『보물섬』이 『정글북』이나 『피터팬』보다 물리적으로 소설의 길이가 길고, 『걸리버여행기』, 『로빈슨크루소』, 『톰소여의모험』보다는 물리적으로 소설의 길이가 짧다는 것을 의미한다. [표 1]과 [표 2]에서 볼 수 있듯이, 각 소설의 토큰 수가 다르므로 이때는 표준화된 TTR을 구하여 비교할 필요가 있다. 각 소설의 STTR을 측정하면, 그 결과 『보물섬』은 81.45고, 『걸리버여행기』는 82.59, 『로빈슨크루소』는 78.51, 『정글북』은 81.01, 『톰소여의모험』은 80.58, 『피터팬』은

81.81이다.

이와 같은 결과는 『보물섬』의 어휘 다양도가 『걸리버 여행기』와 『피터팬』보다는 낮고, 『로빈슨크루소』, 『정글북』, 『톰소여의모험』 보다는 높다는 것을 의미한다. 작가의 어휘 다양도가 높을 수록 작가는 해당 작품에서 다양한 어휘를 사용한다는 것을 의미하고, 그것은 결국 그만큼 독자로서는 그만큼 아는 어휘가 많아야 한다는 것을 의미하며, 결과적으로 모르는 단어가 많을 수 있다는 것을 의미하여 책을 읽기가 어렵다는 것을 의미한다. STTR 결과만 놓고 보면, 『로빈슨크루소』가 물리적인 전체 어휘 수는 많지만 어휘 다양도가 낮아서 이 6개의 책 중에 가장 읽기 쉬운 책이고, 그다음은 『톰소여의모험』, 『정글북』, 『보물섬』, 『피터팬』 순서이고, 『걸리버여행기』는 물리적인 전체 어휘 수도 많고, 어휘 다양도 또한 가장 높아서 이 6개의 소설 중 가장 읽기 어려운 책이라고 할 수 있다. 『보물섬』은 어휘 수나 어휘 다양도 측면에서 이 6개의 소설 중에 중간 정도에 있다고 할 수 있다.

2. 보물섬코퍼스 고빈도 어휘 분석

텍스트에서 어떤 어휘의 빈도가 높다는 것은 글쓴이가 강조하고 싶은 중요한 어휘일 가능성이 크다. 하지만 상황에 따라서는 그 말이 반드시 맞는 것은 아니다. 예를 들어 다음은 보물섬에 나오는 문장이다.

His stories were what frightened people worst of all. Dreadful stories they were - about hanging, and walking the plank, and storms at sea, and the Dry Tortugas, and wild deeds and places on the Spanish Main.

위 두 개의 문장에서 사용된 어휘는 총 36개이다. 즉, 이 문장들에서 사용된 어휘의 빈도수를 측정하면, *and*는 5회, *the*는 3회, *stories*는 2번, *were*는 2번, 나머지 어휘는 모두 1회 사용되었다. 수치상으로 *and*가 5회 사용되었으므로 이 문장들에서 가장 중요한 어휘이어야 한다. 그리고 *the*는 3회 사용되었으므로, *the* 역시 이 문장들에서 두 번째로 중요한 어휘여야 한다. 하지만 실상은 그렇지 않다.

[그림 1]의 위쪽 그림은 보물섬코퍼스의 어휘를 출현

빈도를 기준으로 제시한 것이다. 보물섬코퍼스에서 가장 출현 빈도가 높은 것은 *the*(4,365회)이다. 로빈슨 스티븐슨이 『보물섬』을 집필하며 사용한 총 70,076개 어휘 중 *the*가 차지하는 비중이 6.2%나 되지만, *the*가 『보물섬』에서 매우 중요한 의미가 있다고 판단하기는 어렵다. 출현 빈도 2위인 *and*(2,874회)도 마찬가지다. 따라서 『보물섬』에 사용된 모든 어휘의 기본형(*be*의 변화형 *am, are, be, been, is, am, was, were*의 기본형 *be*로 간주하는 리마)으로 재정렬하면 [그림 1]의 아래쪽 같은 결과가 나온다. [그림 1]의 위쪽과 아래쪽의 차이는 위쪽은 사용된 총 어휘를 바탕으로 어휘리스트를 추출하였으므로 타입이 5,842개이고, 아래쪽은 리마처리를 한 후 추출한 결과이므로 리마 타입이 4,368개로 타입의 수가 줄어든 것이다.

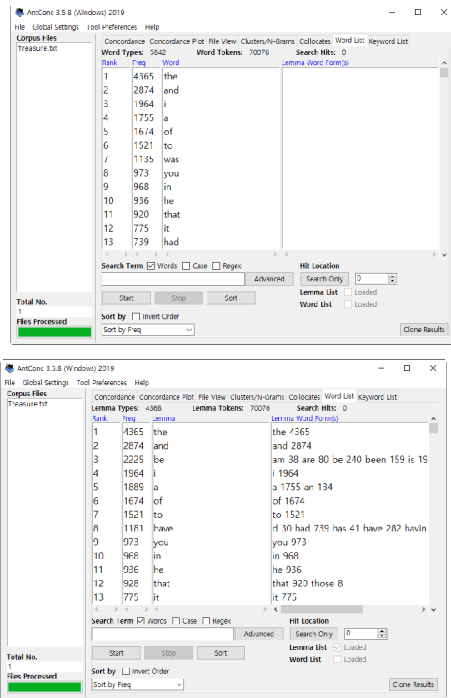


그림 1. 보물섬코퍼스 고빈도 어휘(위)와 리마(아래)

보물섬코퍼스를 리마 처리하여 상위 30위까지 빈도 수대로 결과를 제시하면 [표 3]과 같다.

표 3. 보물섬코퍼스 고빈도 리마 리스트

순위	빈도	리마	순위	빈도	리마	순위	빈도	리마
1	4,365	the	11	936	he	21	486	say
2	2,874	and	12	928	that	22	479	we
3	2,225	be	13	775	it	23	470	but
4	1,964	I	14	651	his	24	433	me
5	1,889	a	15	624	as	25	418	at
6	1,674	of	16	618	my	26	388	this
7	1,521	to	17	616	with	27	367	man
8	1,181	have	18	594	s	28	359	all
9	973	you	19	588	for	29	344	not
10	968	in	20	514	on	30	304	by

〈표 3〉에서 볼 수 있듯이, *the*가 보물섬코퍼스에서 4,365회 사용되었고, 사용빈도 수 1위를 차지하였다. 그 뒤를 이어 *and, be, I, a, of, to, have, you, in* 등이 고빈도 어휘를 차지하고 있고, 30위까지 순위를 보면 대부분이 기능어임을 알 수 있다. 기능어가 대부분인 고빈도 리마 리스트에서 몇 가지 주목할 어휘는 *I*(1,964), *you*(973), *he*(936), *his*(651), *my*(618), *we*(479), *me*(433) 등의 인칭대명사이다. 이러한 고빈도 인칭대명사의 출현은 『보물섬』이 대화체가 많이 사용되는 소설 장르의 특징인 구어체 스타일이라는 것을 계량적으로 보여주는 증거이다. 학술지 논문이나 학술 서적과 같은 문어체 장르의 텍스트에서는 이러한 인칭대명사, 특히 *I*를 찾기 어렵다. 『보물섬』 주어 *I*가 1,964회 사용되어 다른 인칭대명사보다도 월등히 많이 사용되었는데, 그 이유는 『보물섬』은 주인공 짐 호킨스가 보물섬에 관한 이야기를 들려주는 방식으로 진행되기 때문이다. 물론, 대화가 많이 나오는 소설 장르의 특성을 참작하면 *I*의 사용이 늘 짐 호킨스의 발화에서 나오는 것이 아니라 다른 등장인물의 발화에서 나오는 것일 수도 있다는 점도 염두에 둘 필요가 있다.

3. 보물섬코퍼스 키워드 분석

이상에서 살펴본 바와 같이 산술적으로 계산된 고빈도 어휘만으로는 『보물섬』에서 중요한 의미를 담고 있는 어휘를 판단하기 어렵다. 따라서 본 연구에서는 보물섬코퍼스와 BNC 워드리스트와 비교하여 키워드를 추출하는 방식을 적용하였다. 하지만, 이렇게 추출한 키워드 역시 기능어가 상당수 상위권에 포진하고 있었고, 등장인물의 이름과 고유명사 등이 다수 상위권에 포함되

어있었다. 결국, 기능어, 등장인물의 이름, 고유명사 등을 제외한 명사, 동사 위주의 내용어를 필터링하여 추출하게 되었고, [표 4]는 이상과 같은 과정을 통해 도출된 결과의 상위 30개를 키니스를 기준으로 제시한 것이다. 이때 키워드 추출을 위해 사용한 방식은 로그라 이클리후드 통계방식이며, p값은 0.05보다 작고($p < 0.05$) 본페로니 교정한 결과이다.

표 4. 보물섬코퍼스 주요 키워드

어휘	키니스	빈도	어휘	키니스	빈도	어휘	키니스	빈도
captain	1435.42	234	island	299.39	81	aboard	202.75	35
doctor	748.04	172	schooner	299.36	28	cutlass	183.14	15
cried	605.49	103	coracle	294.75	25	dead	180.08	73
ship	579.47	116	lay	279.94	89	dooty	179.03	13
man	489.67	260	anchorage	261.86	29	buccaneers	167.35	16
rum	469.93	57	ashore	254.62	107	cabin	160.58	31
stockade	409.88	39	mutineers	240.13	38	began	153.7	90
treasure	397.93	59	sir	237.03	23	sand	151.95	40
cap	321.1	61	crutch	219.21	24	mate	148.22	33
sea	316.96	110	deck	213.42	41	forecastle	143.04	11

앞서 [표 3]에서 제시한 30개 고빈도 리마로는 『보물섬』의 내용이나 『보물섬』이라는 작품을 특정 지을만한 단서를 찾기가 어려웠다. 하지만 [표 4]에서 제시된 30개 주요 키워드를 보면 이전과 다른 점을 확인할 수 있다. 이 30개의 키워드만으로 『보물섬』의 분위기와 내용을 대략 가늠할 수 있다. 예컨대, 『보물섬』의 내용을 전혀 모르는 사람도 작가가 사용한 70,076 어휘 중 코퍼스 분석을 통해 추출한 0.014%인 *captain*(선장), *doctor*(의사), *ship*(배), *man*(남자), *rum*(럼주), *sea*(바다), *treasure*(보물), *island*(섬), *buccaneers*(해적), *dead*(죽은) 등 10개의 키워드만 보아도, 등장인물의 직업, 럼주와 배가 나오고, 바다, 보물, 섬, 해적, 죽은 것들이 나오는 것으로 보아 보물섬과 연관된 흥미진진한 해적 소설이라는 추측이 가능하다. 작가가 사용한 어휘 70,076 중에서 키니스가 높은 30개 키워드를 보면 『보물섬』의 내용을 조금 더 구체적으로 가늠하는 데 도움이 된다. 이러한 키워드는 『보물섬』을 처음 접하는

아동 독자에게 『보물섬』을 읽기 전에 이 핵심 30개 키워드를 학습하고 난 후에 읽게 하면 『보물섬』의 전체적인 내용을 이해한 상태에서 책을 읽기 때문에 『보물섬』에 관한 이해가 더 빠르고 풍성해질 수 있을 것이다.

『보물섬』이 해상여행과 해적과 소년의 싸움 등 남성 독자를 자극할 만하였기 때문에 남성 독자에게 많은 관심과 인기를 끌 수 있었다고 한다[27]. 영국 아동 모험 소설의 성 역할에 관한 연구를 통해 답론 대부분이 남성에게 의해 지배된다고 주장도 있는데[28], 『보물섬』이 남성주의적인 성향을 갖고 있다는 주장을 뒷받침하는 키워드 *man*(남자), *mate*(남자 동료)를 [표 4]에서 찾을 수 있다. 이러한 키워드를 중심으로 『보물섬』에서 나타난 남성성을 찾아보면, *strong*을 찾을 수 있다. *strong*은 『보물섬』에서 총 22회 사용되었고, 주로 강한 남성을 묘사할 때 사용되었다(예: *five strong and active seamen!*, *his jaws as strong as iron*, *He was very tall and strong*). 또 다른 남성성을 상징하는 *rough* 역시 다양한 상황에서 남성들의 거친 모습이나 성향을 보여주는 의미로 사용된 것으로 나타났다(예: 1) *some of my hands being rough*, 2) *I'm a man that have lived rough*, 3) *They lives rough, and they risk swinging*)

4. 보물섬코퍼스 n-그램 분석

본 연구에서는 『보물섬』에 대한 이해와 작가에 대한 이해를 넓히고자 3-그램과 4-그램 측정을 하였고, [표 5]는 각각의 상위 20개의 결과이다.

표 5. 보물섬코퍼스의 상위 3-그램과 4-그램

순위	빈도	3-그램	빈도	4-그램
1	39	I don't	25	you may lay to
2	33	out of the	19	may lay to that
3	29	one of the	17	and you may lay
4	28	said the captain	13	I'll tell you
5	27	I began to	11	at the same time
6	25	it was a	10	of the spy glass
7	25	may lay to	9	and a bottle of
8	25	you may lay	9	at the same moment
9	23	at the same	9	I don't know
10	23	of the island	9	I don't like
11	22	I could see	8	a bottle of rum
12	22	to be sure	8	ho and a bottle
13	21	the captain's	8	ho ho and a

14	20	and I'll	8	yo ho ho and
15	20	as soon as	7	all of a sudden
16	20	said the doctor	7	as soon as I
17	19	I could not	7	eight pieces of eight
18	19	lay to that	7	of eight pieces of
19	19	that I had	7	pieces of eight pieces
20	18	in the meantime	7	the dead man s

[표 5]에 제시된 결과에서 우선 주목할 것은 주어 /의 사용이다. 전언하였듯이 대화체가 많은 소설에서는 주어는 자주 사용되는 어휘다. 특히 *I could see*는 물리적으로 뭔가를 볼 수 있다는 의미와 함께 문맥 속에서 뭔가를 깨닫게 되었다는 의미로도 사용된다. 예컨대, 1) *I could see at once that we were ruined*, 2) *I could see it was in vain for me*, 3) *I could see the truth of Silver's words* 등에서 볼 수 있듯이 작가는 『보물섬』에서 중간중간에 *I could see* 문구를 사용함으로써 등장인물이 뭔가를 깨닫게 하고, 이후 이를 바탕으로 후속 생각과 행동으로 이어질 수 있도록 장치를 한 것으로 해석할 수 있다. 또한, *I began to* 역시 유사한 의미로 사용된 것으로 보인다. 예를 들어, 1) *I began to believe that I had found an ally*, 2) *And then, of a sudden, I began to comprehend*, 3) *I began to get another thought into my head* 등에서 볼 수 있듯이 등장인물이 *I could see*와 유사하게 뭔가를 깨닫고 새로운 아이디어를 갖게 되고, 이를 바탕으로 새로운 국면을 맞이하고 적극적으로 문제를 해결해 나가기 시작한다.

고빈도 4-그램에서 주목할 것은 작가의 동시에라는 표현이다. 『보물섬』 작가는 동시에라는 표현을 *at the same time*(11회)만 사용하지 않고 *at the same moment*(9회)를 번갈아 사용했는데, 두 표현을 합쳐 20회 사용한 작가는 혼치 않다. 예를 들어 같은 아동 모험 소설로 분류되는 『걸리버여행기』의 작가는 *at the same time*을 『걸리버여행기』에서 17회 사용하였지만, *at the same moment*는 전혀 사용하지 않았다. 『로빈슨크루소』의 작가도 *at the same time*을 『로빈슨크루소』에서 7회 사용한 반면, *at the same moment*는 1회 밖에 사용하지 않았다. 한편, 『정글북』의 작가는 *at the same time*과 *at the same moment*를 『정글북』에서 전혀 사용하지 않았고, 『톰소여의모험』의 작가는 *at the same time*을 『톰소여의모험』에서 1회 사용하

였고, *at the same moment*를 전혀 사용하지 않았고, 『피터팬』의 작가는 *at the same time*을 『피터팬』에서 1회 사용하였고, *at the same moment*를 2회 사용하였다.

『보물섬』에서 나타난 고빈도 4-그램에서는 시간과 관련이 있는 *all of a sudden* 표현도 주목할 만하다. 『보물섬』의 작가는 *all of a sudden*을 『보물섬』에서 7번 반복 사용하였다. 모험 소설 장르의 스토리 전개 상 사건의 극적인 전환을 위해 주로 사용하는 연속어휘구라고 할 수 있다. 예컨대, 『보물섬』에서는 1) *Then all of a sudden there was a tremendous explosion*, 2) *And then, all of a sudden, he stopped, and his jaw dropped*, 3) *And then all of a sudden he was interrupted by a noise* 등 『보물섬』의 작가는 어떤 일이 발생하고 난 후, 갑자기 뭔가가 일어나는 상황을 묘사할 때 자주 사용하였다. 하지만 『톰소여의모험』의 작가가 『톰소여의모험』에서 *all of a sudden*을 불과 2회 사용한 것을 제외하고, 『걸리버여행기』, 『로빈슨크루소』, 『정글북』, 『피터팬』의 작가는 *all of a sudden*을 그들의 작품에서 단 1회도 사용되지 않았다. *All of a sudden*은 보통 *suddenly*로도 사용되므로, 다른 작품에서는 *all of a sudden* 대신 *suddenly*로 사용될 수도 있겠으나, *suddenly* 6개 작품을 모두 검색하여도 『보물섬』 28회, 『걸리버여행기』 9회 『로빈슨크루소』 1회, 『정글북』 1회, 『톰소여의모험』 15회, 『피터팬』 23회 등 『보물섬』이 *suddenly*도 사용빈도가 가장 높았다.

이상과 같은 분석 결과를 바탕으로 갑작스러운 사건 전환, 급작스러운 변화와 같이 생각, 행동, 환경의 갑작스러운 변화를 표현할 때 『보물섬』의 작가는 *all of a sudden* 이나 *suddenly*를 자주 사용하고 선호한다고 볼 수 있고, 특히 *all of a sudden*은 작가 로버트 스티븐슨의 시그니처라고 주장할 수 있는 근거가 될 수도 있다.

V. 결론

본 연구에서는 『보물섬』을 중심으로 코퍼스 분석을

영미 아동 모험 소설의 언어적 특성을 알아보고자 하였다. 프로젝트 구텐베르크를 통해 수집한 『보물섬』의 텍스트 파일을 코퍼스화 하여 보물섬코퍼스를 구축하였고, 어휘, 리마, 키워드, n-그램 등을 측정하였다.

STTR을 측정한 결과 『보물섬』은 다른 영미 아동 모험 소설과 비교하여 중간 정도의 어휘 다양성을 보여졌고, 결과적으로 텍스트의 난도 역시 중간 정도인 것으로 나타났다. 키워드 분석 결과는 고빈도 어휘나 고빈도 리마 분석 결과보다 『보물섬』을 신속하고 한 눈에 파악하는 데 훨씬 효과적이었다. 한편, 키워드 분석 결과를 통해 도출한 남성성과 연관된 결과는 모험 소설이 남성 중심적이라는 직관적 주장에 대해 실제적 증거를 제시하는 역할을 하기도 하였다. 키워드 분석을 통해 도출된 키워드를 중심으로 아동 독자를 위한 『보물섬』 읽기 수업 전, 도중, 후 단계별 수업이 이루어진다면, 아동 독자의 『보물섬』 이해도가 증진할 것으로 생각되며, 본 연구의 키워드 결과를 중심으로 모험 소설 장르에서 관습적으로 남용되는 남성 중심적인 세계관에 관한 비평적 읽기 역량을 배양할 수 있는 기초 자료로 활용될 수 있을 것으로 생각한다.

한편, 본 연구에서는 n-그램을 통한 『보물섬』의 작가인 로버트 스티븐슨의 문체를 분석하였다. 스티븐슨은 시간과 관련한 어휘 사용에 있어서 다른 영미 아동 모험 소설 작가와 달리 뚜렷하게 선호하고, 자주 사용하는 패턴의 연속어휘가 있는 것으로 나타났다. 특히 모험 소설에서 국면 전환이나 갑작스러운 변화에 사용될 수 있는 *all of a sudden*은 분명히 로버트 스티븐슨이 선호하고 자주 사용하는 작가의 표시라고 볼 수 있는 사례였다. 새로운 연구방법은 늘 필요하며[29][30], 이러한 n-그램을 활용한 코퍼스 문체론 연구는 다른 작가나 작품을 분석할 때 유용할 뿐 아니라, 학생들이 존경하는 인물이 쓴 텍스트를 모아 n-그램 분석을 함으로써 존경하는 인물의 글쓰기 패턴을 분석하고, 그 인물처럼 글을 쓰는 연습을 할 수도 있을 것이다.

본 연구에서 시도한 『보물섬』 코퍼스 분석 연구 접근 방법을 통해 아동 독자가 문학 작품의 비평적 읽기 역량을 키울 수 있는 계기가 되기를 바라며, 학습자들이 존경하는 인물처럼 글을 쓸 수 있는 학습 교구재개발을 위한 토대가 되기를 희망한다.

참고 문헌

[1] 김지영, 김정렬, “영어 동화를 활용한 읽기 교육의 효과성 메타분석,” 한국콘텐츠학회논문지, 제16권, 제4호, pp.741-756, 2016.

[2] D. Butts, “The Birth of The Boys’ Story and the Transition from the Robinsonnades to the Adventure Story,” *Revue de littérature comparée*, pp.445-454, 2002.

[3] G. Avery, “Nineteenth Century Children: Heroes and Heroines in English Children’s Stories 1780-1900,” London: Hodder and Stoughton, 1965.

[4] P. Popat and K. Kotadia, “Robert Louis Stevenson’s “Treasure Island” as an adventure novel”, *Asian Journal of Multidimensional Research*, Vol.2, No.6, pp.52-59, 2013.

[5] M. B. Green, “Seven Types of Adventure Tale: An Etiology of a Major Genre,” Penn State Press, pp.1-24, 1991.

[6] 김화선, “청소년소설에 나타난 성장 서사 연구,” *국어교육연구*, 제45권, pp.553-576, 2009.

[7] 나병철, “청소년 환상소설의 문학교육적 의미와 ‘가치의 세계,’” *청람어문교육*, pp.399-430, 2010.

[8] 김번, “18세기 영국의 바다에 비친 마음의 풍경들 : 로빈슨 크루소와 걸리버 여행을 중심으로,” *인문학연구*, 제31권, pp.135-162, 2019.

[9] 이복연, “아동문학 작품에 나타난 주인공의 행동특성에 관한 연구,” *교육학전공*, pp.1-45, 1989.

[10] J. E. L. Cremers, ““He’s More a Man Than Any Pair of Rates of You”: Masculinity in Stevenson’s *Treasure Island* and its Adaptations,” Bachelor Thesis Radboud University Nijmegen English Language and Culture, pp.1-37, 2015.

[11] 이수경, “『피터 팬』에 나타난 네버랜드의 공간적 위상,” *동화와 번역*, 제13권, pp.105-130, 2007.

[12] 선주원, “소설의 환상성 읽기를 통한 소설 교육의 가능성 고찰,” *새국어교육* 제92호, pp.191-221, 2012.

[13] <http://www.icr.or.kr/welcome-2>, 2020.9.1.

[14] 권혁승, 정채관, *코퍼스언어학 입문*, 한국문화사, 2012.

[15] S. Aman and S. Szpakowicz, “Identifying

Expressions of Emotion in Text,” In: *International Conference on Text, Speech and Dialogue*, pp.196-205, 2007.

[16] 하명정, “코퍼스에 기반한 문학텍스트 분석,” *한국콘텐츠학회논문지*, 제13권, 제9호, pp.440-447, 2013.

[17] 박순강, “『프랑켄슈타인』의 감정 어휘 및 표현에 대한 코퍼스 기반 연구,” *영미연구*, 제46권, pp.1-32, 2019.

[18] 정채관, “한국 예비 대학生の 영어 사용 특성 파악을 위한 구어 영어 학습자 코퍼스 구축에 관한 소고: 연구 방법을 중심으로,” *외국어교육연구*, 제34권, 제4호, pp.47-82, 2020.

[19] 정윤나, 정채관, “남북한 중학교 1학년 영어 교과서 어휘 분석 연구: 코퍼스 접근,” *중등영어교육*, 제13권, 제2호, pp.53-70, 2020.

[20] <https://www.laurenceanthony.net/software/antconc>, 2020.7.10

[21] 김태국, 정채관, “초등학교·중학교·고등학교 졸업학력 검정고시 영어 어휘에 관한 코퍼스 기반 연구,” *인문학연구*, 제33권, pp.317-338, 2020.

[22] 권혁승, 정채관, 김재훈, *코퍼스언어학 기초*, 한국문화사, 2018.

[23] 김정렬, “남북한 고등학교 영어교과서 4-gram 언어 비교 분석,” *한국콘텐츠학회논문지*, 제20권, 제7호, pp.274-281, 2020.

[24] 이석재, 정채관, “한국 예비 대학生の 영어 사용 특성 파악을 위한 대규모 공개 영어 학습자 코퍼스 구축 및 분석,” *한국콘텐츠학회논문지*, 제14권, 제11호, pp.1019-1029, 2014.

[25] 정채관, “한국 예비 대학生の 영어 사용 특성 파악을 위한 구어 영어 학습자 코퍼스 구축에 관한 소고: 연구 방법을 중심으로,” *외국어교육연구*, 제34권, 제2호, pp.277-300, 2020.

[26] M. Mahlberg, “Clusters, Key Clusters and Local Textual Functions in Dicken’s,” *Corpora*, Vol.2, No.1, pp.1-31, 2007.

[27] 신경숙, “금융(혁명)시대의 그늘: 로버트 루이 스티븐슨의 『보물섬』과 해적의 재현,” *19세기 영어권 문학*, 제19권, 제1호, pp.67-102, 2015.

[28] E. Poynter, “Talking Time in Children’s Adventure Fiction: Which Gender Controls the Discourse?,” *International Journal of Applied Linguistics and English Literature*, Vol.7, No.5,

pp.87-95, 2018.

[29] 정채관, “『Asia Pacific Journal of Corpus Research』 1권 1호의 연구 동향과 연구 방법에 관한 고찰,” *Asia Pacific Journal of Corpus Research*, 제1권, 제1호, pp.127-132, 2020.

[30] 정채관, 권혁승, “2015 개정 영어과 교육과정에 따른 기본 어휘 및 외래어 목록 변화형 연구: 코퍼스적 접근,” *외국어교육연구*, 제31권, 제2호, pp.175-201, 2017.

저 자 소개

최 은 샘(Eunsaem Choi)

준회원



- 2020년 6월 ~ 현재 : 인천대학교 코퍼스연구소 연구인턴
- 2016년 3월 ~ 현재 : 인천대학교 영어영문학과 학부생

<관심분야> : 코퍼스 문제론

정 채 관(Chae Kwan Jung)

정회원



- 2000년 7월 : University of Birmingham 생산공학·일본어(공학사)
- 2002년 1월 : University of Warwick 경영공학 (이학석사)
- 2011년 2월 : University of Warwick 응용언어·영어교육(교육학박사)

- 2011년 3월 ~ 2011년 7월 : 서울대학교 교수학습개발센터 선임연구원
 - 2012년 9월 ~ 2019년 8월 : 한국교육과정평가원 교육과정교과서본부·교육평가본부 부연구위원
 - 2019년 9월 ~ 현재 : 인천대학교 영어영문학과 조교수
 - 2020년 1월 ~ 현재 : 인천대학교 코퍼스연구소 소장
- <관심분야> : 코퍼스, 특수목적영어, 컴퓨터보조번역, 남북한영어교육