

소설 코퍼스의 계량적 분석과 한국문학 교육용 텍스트 선정

황혜란(홍익대학교한국어교육원)

1. 머리말

문학텍스트는 언어교육을 위한 효과적인 자료이며, 한 언어를 사용하는 사람, 사회, 그리고 문화와 역사 전반을 총체적으로 접할 수 있게 한다.¹ 따라서 문학은 수준 높은 한국어 능력을 기르기 위해서 필요하며, 한국문학전공자에게는 전공지식의 습득과 함께 언어능력의 향상도 함께 도모하게 한다.²

그 동안 해외 한국학 한국문학 전공교육뿐만 아니라 국내의 한국어 교육에서도 체계적인 한국문학 텍스트 선정에 대한 요구와 필요성은 인정이 되었으나 문학텍스트의 선정은 다분히 주관적이고 임의적으로 이루어졌다. 이러한 이유는 문학텍스트의 속성상 텍스트를 선정하는 일이 하나의 요인만을 고려할 수 없을 만큼 복잡하고 정교함을 요하는 난해한 작업이며 대상의 범위인 문학작품의 수도 방대하기 때문이다.

문학텍스트의 선정은 언어적·문화적·문학적 측면이 모두 고려되어 통합적으로 이루어져야 한다.³ 그러나 우선적으로 고려해야 할 부분은 언어적 측면이라고 할 수 있다. 제 2언어 한국어학습자는 일반적인 한국인들과 달리 문학작품을 읽을 때 어휘, 문법, 배경지식의 순으로 어려움을 느낀다.⁴ 언어적 기준을 적용하여 문학텍스트를 선정하고 위계화한

1 문학작품이 외국어교육현장에서 활용되는 이유는 문학작품이 가치 있고 믿을만한 자료이며, 목표언어권의 문화적 풍요함을 담고 있고, 문학작품을 통해 목표 언어의 다양한 언어적 가치를 학습하는 과정에서 인간적인 참여도 가능하기 때문이다. Collie & Slater, 조일제 역, 『Literature in the Language Classroom』, 한국문화사, 1997, 16~19쪽.

2 문학을 활용한 언어활동이 실용적인 차원에서 이루어지므로 문학작품은 도구 교과라는 차원에서 언어활동의 좋은 자료이자 학습자가 목표로 하는 언어의 사회 문화적 이해를 넘어 고급스러운 언어 사용교육으로 나아갈 수 있다. 또한 한국문학에 대한 교육은 문학 능력을 함양하고, 세계문학 속에서 한국문학의 보편성과 특수성에 대한 이해를 도모할 수 있다. 윤여탁, 『외국어로서의 한국문학교육』, 한국문화사, 2007, 142쪽.

3 정해권·신주철은 한국어교육에서 문학교육에 적합한 문학작품의 위계화 방안을 제시하였다. 위계화를 언어적 측면, 문화적 측면, 문학적 측면 등 세 가지 범주로 제시하고, 이 세 측면의 개별 위계화가 교육의 목적에 맞추어 통합되어 쓰일 수 있는 통합적 위계화의 방안을 제시하였다. 정해권·신주철, 「한국어교육에서 문학 작품의 통합적 위계화 방안」, 『한국어교육』, 제 22권 1호, 국제한국어교육학회, 203~221쪽.

4 한국문학을 학습할 때 외국인 학습자는 어휘(57.2%), 문법(23.3%), 배경지식(13%), 기타(0.7%)의 순으로 어려움을 느끼는 것으로 나타났다. 반면, 한국의 고등학생들은 배경지식(34.25%), 글의

다면 학습자의 언어수준에 따라 작품을 선택할 수 있고, 한국문학 전공자는 전공 지식의 습득과 함께 언어능력의 향상도 도모할 수 있다.

언어수준별로 텍스트를 선정하기 위해서는 문학텍스트의 어휘, 문법, 문장, 수사적 표현의 모든 요소들을 고려하여 각 항목을 나누어 계량하고 난이도에 따라 등급을 부여하는 방안이 필요하다. 그리고 각 기준들의 상관관계를 분석하고 가중치를 적용한다면 통합적인 언어기준을 마련할 수 있다. 언어적 기준 중에서도 어휘는 언어 난이도를 결정하는 가장 중요한 요인이다.⁵

따라서 본 연구는 텍스트 선정의 임의성을 배제하고 과학적·객관적 방법으로 텍스트를 선정하기 위해 한국문학 코퍼스에 대한 계량적 접근 방법으로 어휘 수준별 문학텍스트 선정기준을 마련하고자 한다. 어휘 난이도의 계량은 국립국어원의 『한국어교육 어휘 내용 개발』의 어휘목록을 기반으로 어휘 등급을 부여할 것이다. 그리고 이를 한국문학교육에서 많이 다루어지고 있는 단편소설 작품 12편⁶을 선정하여 어휘적 기준을 적용해 보고자 보고자 한다.

더불어 어휘 난이도 계량 시에 제기되는 다음과 같은 문제에 대해서 논의해 보고자 한다. 첫째, 『한국어교육 어휘 내용 개발』의 타당성에 대한 검증, 둘째, 해외 한국문학 전공자를 대상으로 할 때 사투리나 비속어 등의 어휘나 띄어쓰기 표준화 문제 등의 문학텍스트의 표기기준, 셋째, 이독성 공식을 적용하여 어휘 난이도를 계량할 때 어휘의 등급 부여에 관한 문제이다.

2. 어휘로서의 문학텍스트

2.1. 분석 대상 텍스트의 선정

한국학을 전공하는 학생들은 한국인과 동일한 수준으로 대학과정을 마쳐야 한다. 한국의 대학생들은 중·고교 시절 문학작품을 배우고 이러한 학습 배경은 자연스럽게 대학에 까지 이어진다. 따라서 해외한국학 전공자들에게 중등교육과정에서 다루고 있는 작품들을 가르친다면 언어의 습득뿐만 아니라 배경지식의 형성에도 많은 도움이 될 수 있다.

구성 방식(13.26%), 개념(12.71%), 흥미·관심·태도(12.15%), 일상생활과의 관련성(4.42%), 주제(2.76%), 문장(0.55%) 등의 순으로 어려움을 느낀다고 하였다. 조일영·윤창욱, 「한국어교육에서 텍스트 자료수준 평가방안에 관한 연구」, 『한국어문교육』15, 2006, 5쪽.

5 독해가 언어로 표현된 정보로서의 텍스트를 이해하는 과정이라고 할 때, 독해에 동원되는 독자의 지적 능력은 언어 지식의 중심이 될 수밖에 없으며, 언어 지식 중에서 어휘 지식이야말로 핵심적인 능력이 될 수밖에 없다. 김종택, 『국어어휘론』, 탑출판사, 1992. 9~11쪽.

6 각주 14번 참조.

이러한 면을 고려했을 때, 방대한 양의 문학텍스트 중에서 교과서 수록 문학작품은 충분한 교육적 근거를 가지고 있고 합리적인 선택기반이 될 수 있으며, 이러한 문학교육의 효용성에 대해서는 한국 내의 한국어 교육자들도 대체로 동의하는 바이다.⁷

외국어교육에서 문학텍스트를 선정할 때 학습자들의 흥미가 중요한 요인으로 작용하기 때문에 텍스트의 언어적 난이도나 적절성을 강조하는데,⁸ 한국문학교육도 이와 맥락을 같이 한다. 그러나 언어적 적절성에 대한 객관적이고 체계적인 기준의 제시는 미비하다고 할 수 있다. 본 연구는 어휘 난이도의 계량을 통해 문학텍스트의 선정 기준을 마련하는 바, 문학 텍스트를 선정하면서 언어적 요인 이외의 다른 요인들을 고려하여 어휘 분석 대상 텍스트를 선정하고자 한다. 주제, 문학사적 가치, 창작 시기, 장르 등을 고려한 텍스트의 선별 기준은 다음과 같다.⁹

첫째, 주제에 있어 보편성을 가지고 있는 작품이다. 문학은 인간을 탐구하고, 삶의 진실을 구현하고자 한다는 점에서 세계문학으로서의 보편성을 지닌다. 학습자의 한국어 수준이 어떠한 해외한국학 전공자들은 자국의 문학교육을 통해 문학텍스트를 읽고 감상할 만한 배경지식과 능력을 지니고 있을 것이다. 따라서 인간, 가족, 사랑 등의 보편적 주제를 가진 작품은 학습자들에게 좀 더 친숙하게 다가갈 수 있으므로 한국적인 특수성을 지닌 작품보다는 보편성을 지닌 작품을 우선 선정하였다.

둘째, 문학사적으로 그 가치를 인정받는 작품을 선정하였다. 한국학을 전공하는 학생들은 한국 문학의 전체적인 모습과 흐름을 파악해야 한다. 문학사적 지식을 단편적으로 학습하기보다는 작품을 통해 흐름을 파악하는 것이 구체적일 것이다.

셋째, 학습자의 실생활과 가까운 시기의 작품을 선정하였다. 문학텍스트를 교육함에 있어서 학습자의 관심과 흥미는 매우 중요한 요소이다. 학습자가 처한 환경과 시대적으로 거리가 가까운 작품이 어휘나 문법 등의 언어적 요인의 최근 경향이 더 잘 반영되어 있

7 윤여탁, 「한국어교육에서 현대문학 정전 연구」, 『국어교육연구』, 서울대학교 국어교육연구소, 2002, 39~64쪽.

8 Littlewood(1984)는 문학작품을 가르칠 때 학습자의 언어능력과 작품 속의 난이도를 고려해야 하며, 학습자의 흥미와 일상 생활과 관련이 있는 것, 그리고 고어체를 피하고 일상적인 언어와 관련된 문체를 지닌 자료를 선정해야 한다고 하였다. Collie&Slater(1997)도 학습자들의 흥미와 관심, 그리고 반응을 이끌어내기 위해 언어적 난이도를 고려해야 한다고 하였다. Lazar역시 (1993) 학습자의 문화적 배경과 언어숙달도를 중요한 요인으로 꼽고 있다.

9 나정선은 작품 선정의 원리를 교육의 목표와 관련하여 언어표현의 풍부함과 다양성, 한국어의 표본이 될 수 있는 표현을 포함한 작품, 한국인의 사고 방식이 전형적으로 드러난 작품을 선정기준으로 제시하였다. 나정선, 「외국인을 위한 문학 교육 방법 연구」, 단국대박사학위논문, 2008.

신윤경은 문학교육에 있어서 문학 텍스트 선정 기준으로 교육과정과 목표의 고려, 작품 자체, 주제 및 활용성을 제시하고, 12가지의 구체적인 체크리스트를 활용하여 작품을 선정하고자 했으며 교과서 수록 수필과 단편소설을 중심으로 목록을 정리하였다. 신윤경, 「한국어교육을 위한 문학텍스트 선정」, 고려대박사학위논문, 2008.

고, 학습자의 공감을 이끌어내기에도 용이하다.¹⁰

마지막으로, 장르에 있어서는 대상을 단편소설로 제한하였다. 단편 소설은 인물과 사건을 중심으로 서술되며, 삶의 한 단면을 통해 그 시대 사회의 모습도 극명하게 보여준다는 점에서 학습자의 흥미유발이나 언어·문화적인 면의 교육에 있어 효용성이 크다고 할 수 있다. 또한 텍스트의 길이가 짧아 중·장편 소설에 비해 전문을 학습하기에 효과적이다.

위와 같은 기준으로 2007개정 교육과정(1학년)에 실린 단편소설 18편 중에서 12작품을 선정¹¹하여 어휘 분석의 대상텍스트로 삼았다. 중복 수록된 작가의 작품은 한 작품만을 선정하였다.¹² 어휘 분석 대상텍스트는 구효서 「이발소 거울」, 김승옥 「서울, 1964년 겨울」, 김유정 「봄·봄」, 박완서 「배반의 여름」, 윤흥길 「종탑 아래에서」, 이문구 「유자 소전」, 이범선 「표구된 휴지」, 이청준 「눈길」, 이태준 「달밤」, 이효석 「메밀꽃 필 무렵」, 황석영 「아우를 위하여」, 황순원 「물 한 모금」 총 12편이다.¹³

2.2. 대상텍스트 어휘에 대한 계량적 분석

분석 대상으로 선정한 12편의 텍스트의 어휘 난이도를 계량하기 위해서는 각 텍스트 어휘의 기본형을 추출하여 기본형 어휘에 대해 등급을 부여하고, 부여된 어휘 등급을 재처리하여 텍스트 별 어휘 난이도 등급을 선정하는 방법을 사용하였다.

먼저, 12편의 분석대상 텍스트의 어휘목록을 만들기 위해 각 텍스트 전문의 기본형을 추출하였다. 그 결과, 분석텍스트의 총 어휘 수는 44,612이고, 작품당 평균 어휘 수는 3,718이다. 텍스트 별 어휘 수는 <표 1>과 같다.

10 446명의 한국어 학습자를 대상으로 설문조사를 한 결과, 학습자가 선호하는 한국어학습 내용은 직접 활용 효과가 큰 일상적 언어사용 기술이나 현재 한국 사회, 문화에 대한 소개 등인 것으로 나타났다. 강승혜, 「한국어교재개발을 위한 학습자 요구 분석」, 『외국어로서의 한국어교육』, 연세대어학당, 2003년, 260~263쪽.

11 개정 교육과정에 실린 현대소설은 총 49편인데 교과서 별로 중복 수록된 작품을 제외하면 총 37편이다. 이 중에서 단편은 18편(48.6%), 장편은 14편(37.8%), 중편은 5편(13.5%)로 단편소설이 가장 많이 수록되었다.

12 중복 수록 작가의 작품은 김유정의 「금 따는 콩밭」, 「만무방」, 「봄봄」 3편, 이태준의 「꽃나무는 심어 놓고」, 「돌다리」, 「달밤」 3편, 박완서의 「겨울나들이」, 「배반의 여름」 2편, 이청준의 「눈길」, 「선학동 나그네」 2편이다. 이 중에서 「봄·봄」, 「달밤」, 「배반의 여름」, 「눈길」 주제와 작품의 인지도를 고려하여 선정하였다.

13 대상 텍스트 12편을 주제별로 분류해 보면, 분단의 비극-「종탑 아래에서」, 내면의 성장-「배반의 여름」, 「아우를 위하여」, 「이발소 거울」, 소외된 인간의 모습- 「달밤」, 「서울, 1964년 겨울」, 「유자 소전」, 가족의 사랑-「눈길」, 기타 개성적 주제- 「봄·봄」, 「메밀꽃 필 무렵」, 「물 한 모금」, 「표구된 휴지」이다. 전은주, 앞의 논문, 21~27쪽.

<표 1> 텍스트 어휘 수

작품 필드	어휘 수
「구효서_이발소거울」	3746
「김승옥_서울, 1964년 겨울」	4312
「김유정_봄·봄」	2967
「박완서_배반의여름」	3340
「윤흥길_종탑아래에서」	5269
「이문구_유자소전」	8772
「이범선_표구된 휴지」	1099
「이청준_눈길」	5852
「이태준_달밤」	1851
「이효석_메밀꽃필무렵」	1966
「황석영_아우를위하여」	4072
「황순원_물한모금」	1366
총 12편	44,613
평균 어휘 수	3,718

대상 텍스트의 기본형 어휘를 계량한 결과는 위와 같다. 위 결과에서 두드러지는 특징은 텍스트에 따라 어휘 수의 편차가 크다는 것이다. 이문구의 「유자소전」의 경우 어휘 수가 8,772개로 가장 많은 반면 이범선의 「표구된 휴지」는 1,099개로 그 차이가 무려 7,673개로 약 8배에 가까운 큰 편차를 보이며, 평균 어휘 수와 비교해도 약 2.4배의 편차가 있다. 어휘 수가 많다는 것은 작품의 길이가 길다는 것을 의미하는데 이는 어휘 등급 통계처리 시 큰 변수로 작용할 수 있다. 따라서 「유자소전」의 경우 평균 값과의 편차를 줄이기 위해 ‘어휘 수 x 0.5’, 「표구된 휴지」의 경우는 ‘어휘 수 x 3’ 등으로 평균 값 대비 비율을 적용하여 계량하는 방안을 생각해 볼 수 있다. 이렇게 한다면 어휘 수의 차이에서 오는 어휘 계량의 변수를 줄일 수 있을 것이다.

3. 텍스트의 난이도 분석

3.1. 『한국어교육 어휘 내용 개발』 어휘 목록 기반

대상텍스트의 어휘목록에 등급을 부여하기 위해서는 일정한 기준이 필요하다. 본 연구

에서는 국립국어원의 『한국어교육 어휘 내용 개발』1,2단계 어휘 목록을 기반으로 텍스트의 어휘 난이도 수준을 분석하고자 한다.

『한국어교육 어휘 내용 개발』연구는 2002년 국립국어원에서 수행한 한국어 학습용 어휘 선정을 위한 기초 조사인 '현대 국어 사용 빈도 조사'와 2010~2011년까지 실시한 '국제통용 한국어 교육 모형 개발'의 연장선에서 진행된 국가 수준의 한국어 교육용 어휘 내용 개발 연구로 초급 단계 어휘 1,836개, 중급 단계 3,855개의 어휘 목록을 선정하였다. 이 연구는 기존의 어휘 선정 연구가 '빈도' 기반으로 하였던 것과는 달리 어휘의 사용 범위 (range)와 포괄성(coverage), 사용 가능성, 친숙도 등을 모두 고려한 객관적 자료를 근거로 하였고, 사용자 평가나 전문가 평정의 주관적 방법도 고려하여 균형을 갖추었으며, 현장에서의 활용성을 고려했다는 점에서 진일보한 연구라 할 수 있다.¹⁴ 따라서 본 연구에서는 『한국어교육 어휘 내용 개발』 어휘 목록을 문학텍스트에 적용시켜 봄으로써 어휘목록 선정에 대한 타당성을 검증하고자 한다.

분석대상 텍스트에 어휘 등급을 부여하기 이전에 『한국어교육 어휘 내용 개발』 1단계의 초급 어휘 1,836 개는 A, 2단계 중급 어휘 3,855 개는 B, 1,2 단계에 해당하지 않는 어휘는 고급으로 C로 분류한다.¹⁵

3.2. 어휘 표기의 정상화

분석대상 텍스트의 어휘 난이도를 계량하기 이전에 완성된 두 개의 목록의 표기를 살펴 보고자 한다.¹⁶ 『문학텍스트』와 『어휘 내용 개발』의 어휘 목록은 표기에 있어 방식이 다르다. 먼저, 『어휘내용개발』은 '단어'를 표시할 때 동형이의어가 있는 어휘에는 두 자릿수의 아라비아 숫자를 첨자(superscript)하였다. 예를 들면, '가족⁰¹, 가지⁰¹, 가지⁰⁴'와 같다. 반면, 동형이의어가 없는 어휘에는 아무런 표시를 붙이지 않았다. 예를 들면, '가지다, 가짜'와 같은 경우이다. 이는 『표준국어대사전』의 데이터베이스의 방식과 같다. 그리고 『어휘내용개발』에서는 품사 이름을 별도의 필드로 제시하였고, 어휘의 의미 이해를 돕기

14 이 연구는 빈도 분석, 한국어 교육과 국어 교육 분야의 가족 어휘 선정 목록 간의 중복도 분석, 한국어 교재의 텍스트 및 한국어능력시험(TOPIK)텍스트로 구성된 한국어 교육 말뭉치에서의 어휘량 분석, 한국어 학습자 및 교사의 어휘 친숙도 조사, 한국어 교육 전문가의 어휘 평정 등 다양한 어휘 선정 방법을 단계화하여 체계적으로 수행함으로써 어휘 선정 연구의 새로운 모형을 보였다. 국립국어원, 『한국어교육 어휘 내용 개발』, 2014, 174~175쪽.

15 『한국어교육 어휘 내용 개발 체계』3단계 연구는 현재 진행 중이다. 3단계 연구가 완료되면 고급 어휘 목록이 추가로 공개될 예정이다. 3단계 연구가 마무리 된다면 현재, '초급-중급-고급'의 3단계 등급에서 '초급-중급-고급-최고급'의 4단계 등급으로 좀 더 정교한 어휘 난이도 계량이 가능할 것이다.

16 이후 분석대상 텍스트의 어휘 목록은 『문학텍스트』로 『한국어교육 어휘 내용 개발』의 어휘 목록은 『어휘내용개발』로 약칭한다.

위해 길잡이 말을 제시하고 있다. 길잡이 말은 주로 일상생활에서 보편적으로 사용되는 연어나 유의어, 반의어, 상위어 등이다. 예를 들어, ‘가르치다’의 길잡이 말은 ‘한국어를’, ‘강’의 길잡이 말은 ‘바다’와 같은 식이다.

그러나, 『문학텍스트』는 ‘어휘-첨자-품사기호’형식의 문자열로 표시된다. 동형이의어가 없는 어휘에 ‘00’번호를 붙였고, 품사 이름은 김병선¹⁷이 사용하고 있는 두 자릿수의 영문 표기로 바꾸어서, 최종적으로 어휘와 첨자와 품사기호를 묶었다. 품사기호는 ‘21세기 세종계획’의 기호를 반영하되, 간략화한 영문 알파벳 두 글자로 통일하였다. ‘21세기 세종계획’과는 달리 보조동사, 보조형용사는 구분하였다.

『문학텍스트』의 어휘에 등급을 부여하기 위해서는 『어휘내용개발체계』를 기반으로 두 목록을 상호대조 한다. 이를 위해서 두 목록의 표기방식을 통일할 필요가 있다. 목록 간 상호 대조를 통해 어휘 등급을 부여하는 ‘자동 재처리 과정’을 위해 김병선의 방식에 따라 표기를 통일하였다.¹⁸ 이 방식으로 처리한 『문학텍스트』의 어휘 목록에는 『어휘내용개발』의 어휘 등급이 자동적으로 부여된다. 예를 들어 『문학텍스트』의 ‘가다01vv’는 『내용개발체계』의 ‘가다01’에 부여된 ‘초급’ 어휘 등급 A가 자동적으로 부여 되는 것이다.

여기서 한 가지 생각해 볼 문제가 있는데, 어휘 난이도를 계량 하기 위해 텍스트의 표기를 어떻게 할 것인가의 문제이다.

먼저, 문학 작품 속에 자주 등장하는 사투리나 비속어의 문제를 살펴 보겠다. 이문구의 「유자소전」같은 경우 충청도 사투리와 비속어가 자주 등장한다. 충청도 사투리의 경우 작품 속에서 토속적인 정감과 사실성을 주고, 주인공에게 친숙함을 느끼게 하고, 비속어는 작중 인물이 처한 상황과 그 인물이 대변하고 있는 사회의 모습을 효과적으로 표현하며 비판의 대상을 더욱 우스꽝스럽게 보이게 하는 풍자적인 효과를 준다. 그러나 사투리나 비속어는 어휘 난이도를 높이는 요인이기도 하다.

예를 들어, ‘대관절 월매짜리 고기간디 그려?’ 라는 문장에서 ‘월매’는 어휘 등급 부여시 고급인 C 등급으로 처리된다. 그러나 이것을 표준어인 ‘얼마’로 대체했을 경우는 초급인 A로 등급이 부여된다. 이는 비속어의 경우도 마찬가지이다. 예를 들어 ‘딴따라’, ‘꼬랑지’, ‘성깔’, ‘처먹다.’ 등의 어휘들도 그대로 처리했을 경우와 대체어로 치환해서 표기한 경우에 난이도 차이가 발생한다.

17 김병선은 그가 구축한 8,201편의 현대시 코퍼스(Poetry Corpus: KoPoKo)를 기반으로 시 텍스트의 어휘에 5단계의 등급을 부여하고, 시 텍스트의 어휘 등급에 따른 선정방안을 제시하였다. 이는 언어적 기준을 코퍼스에 기반한 과학적이고 객관적인 방법으로 제시했다는 점에서 의의가 있다. 김병선, 「한국 현대시의 언어 등급 부여 방안」, 『제1차 서태평양 한국어교육 및 한국학 국제학술회의』, 2011, 363~382쪽.

18 김병선(2011)의 표기 방식은 다음과 같다. 일반명사(ng), 고유명사(nm), 의존명사(nb), 대명사(np), 수사(nr), 동사(vv), 형용사(va), 보조동사(vx), 보조형용사(vz), 관형사(mm), 부사(ma), 감탄사(ic), 격조사(jk), 어미(ed), 접두사(xp), 접미사(xs), 어근(xr), 기호(sb), 분석불능(un)이다.

이것은 비단 사투리의 문제만이 아니다. 1933년 ‘한글 맞춤법 통일안’이 나오기 이전의 작품은 외국어 학습자에게 익숙하지 않은 옛말로 표기되었다. ‘한글 맞춤법 통일안’은 몇 차례 수정을 거쳐 1948년에 공식적으로 적용되었기 때문에 1933년 이후에 창작된 작품들도 이를 바로 적용하지 않은 경우가 많았다. 맞춤법 통일안이 문학작품에 본격적으로 적용된 것은 50년대부터라고 할 수 있다. 따라서 한국어 학습자에게 50년대 이전의 문학작품을 교육할 때 텍스트를 옛말 그대로 표기할 것인지 현대어로 치환하여 표기할 것인지를 문제가 제기된다.

외국어로서의 한국문학교육은 학습자에 따라 요구가 달라지는데, 일반목적 학습자인가 학문목적 학습자인가에 따라 문학을 학습하는 목적이 다르기 때문이다. 이러한 차이는 무엇을 중심으로 문학을 가르칠 것인가의 문제와 관련된다.¹⁹ 일반목적 학습자를 대상으로는 언어중심의 교수를 할 때는 현대어로 모두 표기한 것이 더 효과적일 수 있으며, 학문목적 학습자를 대상으로 문학중심의 교수를 할 때는 가능한 원본에 충실한 것이 바람직할 것이다.

해외에서 한국학을 전공하는 학생들은 전공지식의 습득에 대한 열의와 수학 능력을 가지고 있다. 그러나 한국어의 수준은 개인 편차가 많고, 한국에서 유학하는 학생들에 비해 전반적으로 낮을 수밖에 없다. 따라서 한국문학을 교육함에 있어 작품의 언어적 난이도에 따라 등급을 부여하고 선별·교수하는 전략과 함께 한국 문학의 본연의 모습을 교수하는 전략도 필요하다. 그러므로, 해외한국학 전공자를 대상으로 하는 문학텍스트는 현대어화한 표준판이어야 한다. ‘현대어 표준판’은 대상 텍스트를 작가 생존 시 출판한 텍스트를 기준으로 하고 정서법은 현대어로 바꾸어 쓰는 것이다. 현대어는 『표준국어대사전』에 등록된 어휘를 기준으로 한다. 다만, 작가가 의도적으로 사용한 사투리나 문학적 표현은 그대로 받아들이는 것을 원칙으로 한다.

4. 어휘 난이도의 계량²⁰

어휘의 난이도를 결정하기 위해서는 판단 기준이 필요한데 어휘 목록이나 코퍼스가 될 것이다. 이는 각각 『어휘내용개발』과 『문학텍스트』에 해당한다.

현대어 표준판으로 목록화한 『문학텍스트』는 『어휘내용개발』의 어휘목록과 상호 대조하였다. 그리고 이를 재처리했을 때 자동적으로 A-B-C 3단계로 어휘에 등급이 부여되었다.

19 문학텍스트는 외국어교육에서 문학을 통한 언어, 문화에 대한 학습, 그리고 학습자의 체험 표현에 유용하다. 한국문학교육도 이와 맥락을 같이 하여 문학을 통한 한국어 의사소통교육, 한국의 사회·문화 교육, 한국학 차원의 한국문학 교육은 언어모델, 문화모델, 문학모델 또는 인간성장 모델의 세 가지 형태로 보편화 되었다. 류종열, 「외국어로서의 한국문학교육 현황과 과제」, 『한중인문학연구』 35, 2009, 39~40쪽.

20 관련 데이터 분석 자료는 별도 첨부한다.

이렇게 등급이 부여된 어휘를 기반으로 어휘 난이도를 계량한다. 『문학텍스트』의 어휘 난이도를 계량하기 위해 Dale-Chall 이독성 공식²¹을 변형한 공식을 적용하였다.²²

『어휘내용개발』은 기존의 어휘 목록들이 가지고 있던 단점을 보완하고, 한국어교육현장에 맞게 기존 연구성과를 총 망라하여 어휘 목록을 재조정했기 때문에 기존의 목록들이 가지고 있던 한계점에서 진일보한 연구라 할 수 있다. 따라서, 『어휘내용개발』를 기반으로 계량하는 만큼 텍스트의 어휘 난이도의 객관성은 자연스럽게 더 높아진다고 할 수 있다.

어휘의 난이도를 계량하는 방법은 다음과 같다. 어휘의 등급을 초급은 A, 중급은 B, 그 외는 C 등급으로 정하고 각 등급별 단어의 수를 계량한 후 총 종형²³ 대비 각 등급 단어의 비율을 계산한 뒤 가중치를 두어 텍스트의 어휘 난이도 지수를 산출한다. 공식은 다음과 같다.

$$\text{어휘 지수} = (n / A\text{의 } n \times 10) + (n / B\text{의 } n \times 20) + (n / C\text{의 } n \times 30)$$

A=초급 B=중급 C=고급 n=종형개수

어휘는 다음과 같은 원칙으로 계량하였다.

첫째, 모든 품사를 계량한다. 『어휘내용개발』 목록에는 품사가 별도의 필드로 분류되어 있기 때문에 모든 품사를 계량할 수 있다. 이는 동사, 형용사, 명사에 한정하여 종형을 계량하는 기존 연구²⁴와 비교할 때 보다 정교하게 어휘 난이도를 측정할 수 있다.

둘째, 고유명사의 경우 인명, 지명 등은 모두 하나로 계량한다.

셋째, 합성명사의 경우 『표준국어대사전』에 등재된 것을 원칙으로 계량한다. 목록에 없는 합성 명사는 사전에 별개로 나오면 각 구성요소를 각각 나누어 계량하고, 별개로 나

21 Dale-Chall RGS(Reading Grade Score)=(0.1579*DS)+(0.0496*ASL)+3.6365

DS=Dale 단어 목록표에 없는 비율(어려운 단어 수/단어 수), ASL=평균 문장 길이(단어 수/문장 수)

22 구민지는 읽기 텍스트의 어휘지수를 계량하기 위해 Dale-Chall 공식을 변형하여 김종섭(2011) 『국제통용한국어교육 표준모형개발』의 목록을 기준으로 각 단어의 등급을 A~C 등급까지 정하고 목록에 없는 단어를 D등급으로 처리하여 각 등급별 단어의 수를 계량하였다. 그리고 총 종형 대비 각 등급 단어의 비율을 계산한 뒤 가중치를 두어 텍스트의 어휘 난이도 지수를 산출하여 고급 읽기 텍스트에 적용시켜 검증하였다. 구민지, 「한국어 읽기 교육을 위한 텍스트 난이도 측정법 연구」, 카톨릭대박사학위논문, 2013.

23 ‘개형’은 token, ‘종형’은 type을 나타내며 개별 단어와 종류별 단어를 뜻한다. 예를 들어 “나는 어제 백화점에서 옷도 사고 신발도 샀다.”라는 문장에서 모든 단어를 계량할 때, 개형은 ‘나, 어제, 백화점, 옷, 사다, 신발, 사다’의 7개이고, 종형은 ‘나, 어제, 백화점, 옷, 사다, 신발’의 6개이다.

24 구민지(2013)의 연구에서는 동사, 형용사, 명사에 한정하여 종형을 계량하였다.

오지 않으면 하나의 명사로 계량한다.

넷째, 외래어는 일반명사와 같은 방식으로 계량한다.

다섯째, 다의어의 경우는 사전을 기준으로 별도로 계량한다.

위와 같은 방법으로 『문학텍스트』의 텍스트 별 어휘 난이도를 계량한 후 어휘 수준별로 작품을 선정하고 위계화하였다.

5. 맺음말

지금까지 소설코퍼스에 대한 계량적 분석을 기반으로 하는 한국문학텍스트의 어휘수준별 선정방안에 대해 논의를 전개하였다.

과학적이고 객관적인 방법으로 언어수준별로 텍스트를 선정하기 위해서는 문학텍스트의 어휘, 문법, 문장, 수사적 표현의 모든 요소들을 고려하여 각 항목을 나누어 계량하고 난이도에 따라 등급을 부여하는 방안이 필요하다. 그리고 각 기준들의 상관관계를 분석하고 가중치를 적용한다면 통합적인 언어기준을 마련할 수 있다. 언어적 기준 중에서도 어휘는 언어 난이도를 결정하는 가장 중요한 요인이다.

어휘 수준별 텍스트를 선정하기 위해서 국립국어원의 『한국어교육 어휘 내용 개발』의 어휘목록을 기반으로 소설코퍼스 어휘에 A, B, C 세 단계로 등급을 부여하고, 이를 대상텍스트로 선정한 단편 소설 12편에 적용한 후 대상작품을 어휘 수준별로 목록화하였다.

어휘 난이도를 계량하는 과정은 다음과 같다. 첫째, 분석대상이 되는 텍스트를 선정하였다. 둘째, 분석 대상으로 선정한 단편소설 어휘의 기본형을 추출하여 목록화하였다. 셋째, 국립국어원의 『한국어교육 어휘 내용 개발』과 대상텍스트 목록의 표기를 통일한 후 재처리를 통해 텍스트 어휘에 자동적으로 어휘 등급을 부여하였다. 넷째, 등급화한 문학텍스트 어휘 목록을 기반으로 변형된 이독성 공식에 적용하여 어휘 난이도를 측정하였다. 다섯째, 어휘 난이도에 따라 대상작품을 위계화하였다.

더불어 어휘 수준별 텍스트를 선정하는 과정에서는 『한국어교육 어휘내용개발』의 타당성에 대한 검증에 관한 문제, 해외 한국문학 전공자를 대상으로 할 때 사투리나 비속어 등의 어휘나 띄어쓰기의 표기 기준에 대하여 문제를 제기해 보았다.

본 연구는 텍스트의 어휘 난이도를 코퍼스에 기반하여 계량하는 만큼 한국문학텍스트 선정의 위계화를 위한 과학적이고 객관적인 기준이 될 것으로 기대한다.

<< 참고 문헌 >>

- 국립국어원, 『한국어교육 어휘내용개발』1단계 보고서
국립국어원, 『한국어교육 어휘내용개발』2단계 보고서
- 구인환 외, 『문학교육론』, 삼지원, 2007.
김종택, 『국어어휘론』, 탑 출판사, 1992.
신주철, 『한국어교육에서 한국문학교육의 이론과 실제』, 커뮤니케이션북스, 2006.
윤여탁, 『외국어로서의 한국문학교육』, 한국문화사, 2007.
Collie&Slater, 조일제 역, 『Literature in the Language Classroom』, 한국문화사, 1997.
- 강승혜, 「한국어교재개발을 위한 학습자 요구 분석」, 『외국어로서의 한국어교육』, 연세대어학당, 2003년.
구민지, 「한국어 읽기 교육을 위한 텍스트 난이도 측정법 연구」, 카톨릭대박사학위논문, 2013.
김병선, 「한국 현대시의 언어 등급 부여 방안」, 『제1차 서태평양 한국어교육 및 한국학 국제학술 회의』, 2011.
나정선, 「외국인을 위한 문학교육 방법 연구」, 단국대박사학위논문, 2008.
류종열, 「외국어로서의 한국문학교육 현황과 과제」, 『한중인문학연구』35, 2009.
박청, 「외국인을 위한 한국 소설교육방안 연구」, 이화여대석사학위논문, 2002.
소미, 「한국문학 학습자를 위한 현대어 표준판 시 텍스트 구성 연구」 한국학중앙연구원석사학위논문, 2012.
신윤경, 「한국어교육을 위한 문학텍스트 선정」, 고려대박사학위논문, 2008.
윤여탁, 「한국어교육에서 현대문학 정전연구」, 『국어교육연구』, 서울대학교 국어교육연구소, 2002.
윤영, 「외국인을 위한 한국 소설교육방안」, 이화여대석사학위논문, 1999.
전은주, 「고등학교 국어교과서 현황에 대한 연구」, 고려대석사학위논문, 2012.
조일영·윤창욱, 「한국어교육에서 텍스트 자료수준 평가방안에 관한 연구」, 『한국어문교육』15, 2006
주은정, 「문학텍스트를 활용한 한국어교육 연구」, 경희대석사학위논문, 2002.
최지현, 「영어권 한국어 교재 편찬에 활용되는 한국문학의 범위와 과제」, 『외국어로서의 한국어교육』14집, 서울대학교 국어교육연구소, 2004.
홍서연, 「외국어로서의 한국어교육을 위한 문학텍스트 분류 방안」, 아주대석사학위논문, 2001.
황인교, 「외국어로서의 한국문학교육의 가능태」, 『외국어로서의 국어교육』25·26집, 연세대학교 한국어학당, 2001.
황인교·김성숙·박연경, 「집중적인 한국어교육과정의 문학교육」 『외국어로서의 한국어교육』29집, 연세대학교한국어학당, 2004.