

데이터를 활용한 한중소설 비교 연구

- 이태준의 <농군>과 리훤잉의 <만보산>을 대상으로 -

이택* · 류인태**

차례

1. 머리말: 데이터 기반의 근현대 한국문학 연구 동향
2. 연구대상: 이태준의 <농군>과 리훤잉의 <만보산>
3. 연구방법: 시맨틱 데이터 편찬 및 그래프 시각화
4. 연구결과: 지식그래프를 활용한 작품 비교
5. 맺음말: 디지털 환경에서의 문학작품 연구에 대한 제언

<국문초록>

본 연구는 1931년 만주에서 일어난 만보산 사건을 소재로 한 두 소설 이태준의 <농군>과 리훤잉(李輝英)의 <만보산>을 대상으로 시맨틱 데이터를 구축하고, 해당 데이터를 바탕으로 지식 그래프를 구현함으로써, 시맨틱 데이터 편찬 방법론을 활용한 소설 연구의 가능성을 모색하고자 하였다. 만보산 사건은 한중일 세 나라가 모두 얽힌 역사적 사건으로서, 해당 사건을 소재로 한 한중일 소설은 그동안 각 민족·국가의 입장을 반영하고 있다는 특징으로 인해 꾸준히 논의의 대상이 되어 왔다. 한편으로 그동안 비교학적 차원에서 작품을 독해하는 연구가 다수 시도되었음에도, 정성적 관점에 입각한 결론적 요지 외에 그것들 사이에 얽힌 문예적 차이를 섬세하게 짚어내는 연구는 없었다. 작품 전체의 지형을 비교하여 보기 위해서는 연역적 성격의 규준이 필요하고, 그에 근거한 귀납적 차원의 검토가 함께 이루어져야 하는데, 일반적인 텍스트 읽기로는 그와 같은 방식의 접근이 어렵기 때문이다. 본 연구에서 차용한 시맨틱

* 성균관대학교 국어국문학과 석사과정 (제1저자)

** 성균관대학교 국어국문학과 박사후연구원 (교신저자)

데이터는 Triple(S-V-O) 형식의 데이터 편찬을 준용함으로써, 인문 지식이 지닌 다채로운 스펙트럼을 정교하게 기술하는 데 초점을 둔다. 이와 같은 접근은 소위 디지털 인문학으로 표방되는 데이터 기반의 비선형적(non-linear) 읽기 방법론 가운데 하나로서, 텍스트를 독해하는 과정에서 연역적 기준의 정립과 귀납적 검토의 실행을 가능하게 한다. 최근 데이터 분석 방법을 활용한 작품 연구가 증가하고 있는 상황에서, 시맨틱 데이터 편찬 및 지식그래프 구현을 통한 소설 읽기의 실험적 모색으로서 본 연구가 던지는 시사점이 있을 것이다.

【주제어】 근대소설, 만보산 사건, 이태준, 李輝英, <농군>, <萬寶山>, 디지털 인문학, 시맨틱 데이터

1. 머리말: 데이터 기반의 근현대 한국문학 연구 동향

최근 몇 년 사이 디지털 인문학을 향한 관심이 증가함에 따라 한국문학 연구에서도 디지털 인문학 방법론을 활용한 연구가 꾸준히 발표되고 있다. 고전문학 분야와 근현대 문학 분야를 포괄해 다채로운 문제의식에 입각한 연구가 이루어지고 있으며, 그로부터 산출되는 결과는 기존의 전통적 문학 연구의 그것과는 다르긴 하나 나름의 시사점을 제시한다는 논의가 조금씩 확대되고 있다. 기존 문학 연구에서 중시해온 질적 문제의식에서 벗어나 대규모 데이터를 대상으로 한 양적 분석과 거시적 패턴의 확인은, 분명 이전과는 다른 맥락에서의 연구에 대한 문제의식을 제시하고 있기 때문이다.

한편으로 그와 같은 대규모 데이터 처리 방식에 입각한 디지털 인문학 방법론은 문학사 연구나 문학연구사 정리¹⁾, 문학 환경이나 조류의 시대

1) 대규모 데이터를 구축·분석해서 문학연구사를 검토한 최근의 연구로 ‘김병준·천정환, 『박사학위 논문(2000~2019) 데이터 분석을 통해 본 한국 현대문학 연구의 변화와 전망』, 『상허학보』 제60권, 상허학회, 2020, 443~517면.’을 거론할 수 있다.

적 추이²⁾와 문학 어휘의 변천³⁾ 그리고 잡지나 신문과 같은 대규모 문학 자료를 대상으로 한 분석⁴⁾ 연구에 있어서는 유효할 수 있겠으나, 전통적인 문학 연구에서 중시되어 온 작가 또는 작품을 대상으로 한 질적 연구에 있어서 어떠한 도움을 줄 수 있는지의 질문 또한 제기되고 있다. 소위 대규모 데이터 기반의 ‘멀리서 읽기(distant reading)’가 질적 연구로서의 ‘가까이 읽기(close reading)’를 보완하는 연구 형식이라는 담론은 오랜 기간 거론되어 왔지만, ‘멀리서 읽기’를 표방하는 연구들이 실제 그에 대해 얼마나 유효한 답을 내놓고 있는지의 문제 제기는, 여전히 풀리지 않고 있는 이 방면의 문제의식을 드러내는 징후라 할 수 있다.⁵⁾

이러한 시각은 ‘멀리서 읽기’의 유효함을 부정하는 것이라기보다도 ‘멀리서 읽기’와 ‘가까이 읽기’의 연결 가능성을 찾지 못한 상황에서 제기되

-
- 2) 문학사회의 관점에서 특정 시기의 문학 환경을 데이터로 분석한 최근의 연구로 ‘이재연 외, 『공학으로 인문학 읽기』, 세창출판사, 2021.’에 수록된 이재연의 ‘『문학 동인의 시대』 전후: 한국 작가 네트워크의 통시적 분석 1917-1927.’을 거론할 수 있다. 해당 글은 2017년 7월 ‘Before and After the “Age of Literary Coteries”: A Diachronic Analysis of the Writer Network in Korea, 1917-1927’라는 제목으로 KOREAN NATL COMMISSION UNESCO에서 간행하는 『KOREA JOURNAL』 52에 수록된 적이 있으며, 해당 연구는 2014년 발표된 동일 저자의 논문 ‘이재연, 『작가, 매체, 네트워크 -1920년대 소설계의 거시적 조망을 위한 시론-』, 『사이(SAI)』 제17호, 국제한국문학문화학회, 2014, 257~301면.’에서 구축한 데이터를 활용하였다.
 - 3) 대규모 데이터를 이용해 특정 문학 어휘의 변천을 조망한 최근의 연구로 ‘서재현·김병준·김민우·박소정, 『멀리서 읽는 “우리” — Word2Vec, N-gram을 이용한 근대 소설 텍스트 분석』, 『대동문화연구』 제115집, 성균관대학교 대동문화연구원, 2020, 349~386면.’을 거론할 수 있다.
 - 4) 특정 문학잡지를 대상으로 대규모 텍스트 분석을 시도한 최근 연구로 ‘조은경·드프레메리웨인, 『『별건곤』의 텍스트 근대성 -데이터과학적 언어분석』, 『문화와 융합』 제43권10호, 한국문화융합학회, 2021, 35~54면.’을 거론할 수 있다.
 - 5) 근현대 한국문학 연구에 있어서 ‘멀리서 읽기(distant reading)’에 입각한 연구의 흐름과 최근 연구의 현황에 관해서는 ‘이재연·정유경, 『국문학 내 문학사회학과 멀리서 읽기 -새로운 검열연구를 위한 길마중』, 『대동문화연구』 제111집, 성균관대학교 대동문화연구원, 2020, 295~337면.’의 5번째 장 ‘근현대 한국 문학의 멀리서 읽기’(314~320면) 내용을 참고할 수 있다.

는 답답한 토로에 가깝다. 전통적 문학 연구로서 작가와 작품에 대한 질적 접근은 여전히 중요한 연구 주제에 해당하며, 그것을 효과적으로 탐구하기 위한 연장선상에서 ‘멀리서 읽기’의 가능성을 찾고자 하는 연구자들의 질문 제기는 자연스러운 것이다. 이런 상황에서 대규모 데이터 처리를 전제로 한 ‘멀리서 읽기’가 아니라, ‘가까이 읽기’의 차원에서 데이터 구축 및 처리 방법론을 문학연구에 활용할 수 없는가에 대한 질문이 제기될 수 있다. 데이터 기반의 ‘멀리서 읽기’를 전제로 ‘가까이 읽기’ 또한 데이터로 실행될 수 있을 때, 이른바 두 맥락에 공통적으로 걸친 데이터의 포괄적 자장(磁場)으로서 두 방면의 연구가 교응할 가능성을 모색해볼 수 있을 것이기 때문이다.

이와 관련해 최근에 발표된 ‘전성규·허예슬·최주찬(2021)’⁶⁾은 그러한 문제의식의 연장선상에서 바라볼 수 있는 사례이다. 해당 연구는 신소설 작가인 이인직의 대표작으로 거론되는 <혈의누>, <귀의성>, <치악산>, <은세계> 등의 소설을 대상으로, ‘공포’와 ‘슬픔’의 감정과 연계 가능한 여러 텍스트 데이터를 구축·분석·시각화하는 작업을 수행하였다. 데이터를 매개로 한 계량적 분석이 아니라 정성적 독해에 기반을 둔 데이터 프로세싱을 연구 과정에 접목하였다는 점이 기존 연구와는 다른 차별점에 해당한다. 해당 연구의 결론부에 서술된 아래의 내용은 유의미한 시사점을 던져준다.

이인직·신소설·근대이행기 문학을 이해하는 오랜 규칙을 적극적으로 상대화하면서, 텍스트에 대한 계량적·정성적 방법론을 미약하나마 적용해 보았다. 그리고 그렇게 얻어진 양적 결과값들을 통해 작품을 재독하며 이인직 소설이 남겨 놓은 흔적을 탐지해보았다. 데이터의 단순성이나 자의성 등

6) 전성규·허예슬·최주찬, 『이인직 소설의 감정론: 시각화와 계량화를 기반으로』, 『민족문학사연구』 제76권, 민족문학사연구소, 2021, 135~189면.

의 여러 한계를 갖고 있지만 이인직 텍스트의 특성을 찾고 그 특성을 표현할 수 있는 데이터를 만드는 과정은 여러 단계의 사고나 시행착오를 요구한다. 이인직 텍스트를 읽고 귀납적으로 도출된 데이터이기 때문에 이 데이터가 다른 신소설 작가의 작품과 연계될 수 있을지도 고민해 볼 필요가 있다. 범위가 큰 데이터를 편찬할 때에는 다른 특성들을 갖고 있는 작품군을 통합할 수 있는 데이터 구조화가 요구되는데 이러한 데이터라면 보다 구조적인 통찰이 가능할 수 있겠으나 다른 한편으로 일반론에 머무를 수 있다. 작품군이나 작가군 단위에서 정성적인 연구방법론으로 접근하는 이러한 시도가 향후 신소설이라는 보다 넓은 범주를 데이터화하는 작업에 있어 신소설의 특징적 지점을 데이터화하는 데에 기여할 수 있는 단계적 시도가 될 것이라고 본다.⁷⁾

‘이인직 텍스트의 특성을 찾고 그 특성을 표현할 수 있는 데이터를 만드는 과정은 여러 단계의 사고나 시행착오를 요구’한다든지, ‘이 데이터가 다른 신소설 작가의 작품과 연계될 수 있을지도 고민해 볼 필요가 있다’든지, ‘작품군이나 작가군 단위에서 정성적인 연구방법론으로 접근하는 이러한 시도가 향후 신소설이라는 보다 넓은 범주를 데이터화하는 작업에 있어 신소설의 특징적 지점을 데이터화하는 데에 기여할 수 있을 것’이라든지, 이러한 서술은 모두 ‘가까이 읽기’로서 데이터 프로세싱 기반의 문학 연구가 지닐 수 있는 학술적 의미에 대해 적극적으로 검토한 결과에서 비롯된 것이다. ‘멀리서 읽기’와는 다른 데이터 기반의 작품과 작가 연구가, 정성적 차원의 ‘가까이 읽기’를 토대로 이루어질 수 있음을 은연중 강조하는 것으로 이해할 수 있다.⁸⁾

7) 위의 글, 186면.

8) 작가를 대상으로 ‘가까이 읽기’ 차원의 데이터 구축 및 접근 방안을 모색한 최근의 연구 사례로서 ‘장문석·류인태, 『디지털 인문학과 한국문학 연구(1) : 작가 연구를 위한 시맨틱 데이터베이스 설계』, 『민족문화사연구』 제75권, 민족문화사연구소, 2021, 347~426면’을 거론할 수 있다. 해당 연구는 연속 기획의 일환 가운데 발표된 첫 번째 결과물 이기에, 추후 나올 연구 내용을 함께 살펴볼 필요가 있겠다.

본 연구는 기존 아날로그 환경에서의 ‘가까이 읽기’ 개념이 아닌 데이터 기반의 ‘꼼꼼히 읽기(micro reading)’로서 시맨틱 데이터 편찬 및 그 래프 시각화를 통한 근현대 소설 읽기가 어떻게 가능한지에 대해 실험적 모색을 시도해보고자 한다. 계량적 독해가 아니라 정성적 독해를 위한 능동적 방법론으로서, 문학 데이터 구축·시각화 과정을 이해하는 데 유의미한 참고가 될 수 있을 것이다.

2. 연구대상: 이태준의 〈농군〉과 리휘영(李輝英)의 〈만보산〉

‘만보산 사건’은 1931년 5월부터 7월 사이 당시 중국의 장춘현(長春縣) 근교 삼성보(三姓堡)에서 조선 이주민과 중국 지역민 사이에 일어난 유혈 사태로, 한·중·일 세 나라가 휘말린 큰 규모의 사건이다. 황무지에 수로를 파 수전(水田)으로 만드는 공사 과정에서 중국 지역민과 조선 이주민의 의견이 서로 대립해 폭력 충돌로 번지게 되는데, 7월 1일 일본 경찰이 개입함에 따라 유혈 사태는 끝이 난다.

당시 조선일보 장춘 지국장이던 김이삼(金利三)이 이에 대해 작성한 과장된 보도가 7월 2일 이후 발행된 『조선일보』 사회면을 뒤덮고 호외까지 발행되자, 이 사건은 중국민이 조선 이주민을 습격해 큰 피해를 준 것으로 알려지면서 예기치 않은 배화(排華)-학살(虐殺) 사건을 조선에서 촉발하였다. 7월 2일부터 30일 사이에 조선에서 중국인을 대상으로 벌어진 습격과 폭행 사건은 대략 30여 건 이상으로 추정되며, 이에 따른 전체 사망자 수는 127명, 부상자 수가 393명, 재산 손실이 250만원에 달하는 것으로 확인된다.⁹⁾ 즉 ‘만보산 사건’은, 협의적으로는 5월부터 7월 사이

9) 박영석, 『萬寶山事件研究 : 日帝 大陸侵略政策의 一環으로서의』, 아세아문화사, 1978, 98~101면.

만주에서 일어난 조선 이주민과 중국 지역민 사이의 충돌 사건을 의미하지만, 광의적으로는 충돌 사건 이후 『조선일보』의 오보로 인해 조선에서 일어난 일련의 배화-학살 사건까지 포함한다.

당시 만주 지역의 정치·사회·문화적 상황은 단순하지가 않았다. 만주사변에 불과 2~3달 앞서 일어난 만보산 사건은 당시 만주의 복잡한 지정학적 배경을 내포하고 있기에, 이 사건을 기반으로 창작된 문학 작품들 또한 실타래처럼 얽힌 당시 만주의 현실을 다층적으로 묘사하고 있다. 특히 이 사건을 소재로 삼은 한국, 중국, 일본 세 나라의 소설은, 각 작가의 민족적 배경과 그에 따른 현실 인식의 차이에 따라, 동일한 소재를 차용하고 있음에도 불구하고 그 내용은 서로 다른 특징을 드러낸다.

예컨대 한국의 경우 이태준의 <농군>, 안수길의 <벼>, 장혁주의 <개간>, 중국은 리훤잉(李輝英)의 <만보산(萬寶山)>, 일본의 경우 이토 에이노스케(伊藤永之介)의 <만보산(萬寶山)> 등이 그에 해당하는 대표적 작품들이다. 동일한 역사적 사건을 다룬다 하더라도 ‘기억의 터’로 기능한 만보산 사건¹⁰⁾이라면, 각기 다른 민족·국가적 배경에 있는 각 작가의 창작 의도가 작품에 결정적 영향을 끼쳤을 가능성이 크고, 이로 인해 실제 사건의 각색이나 허구적 요소가 작품 속에 안배되었을 가능성 또한 무시할 수 없다.

이 가운데 이태준의 <농군>은 그동안 여러 해석과 평가를 낳아 왔다.

10) 이와 관련해 다음의 서술을 참고할 수 있다. “만보산 사건은 동아시아 각국이 각기 다른 방식으로 전유해온 기억의 터였다. 만보산 사건은 동아시아 각국의 내부에서 각기 다른 방식으로 망각되거나 파편화되어 기억되어 왔던 것이다. 동아시아가 공유하는 공동의 기억을 환기하는 것도 새로운 ‘동아시아 의식’ 나아가 ‘동아시아 정체성’을 형성하는 데서 중요하고 또한 필요할 것이지만, 지배/피지배, 침략/피침략의 상호 배치되고 갈등하는 기억의 터를 망각으로부터 환기하고 새로운 방식으로 기억하는 것도 반드시 필요한 일이라 하지 않을 수 없다.” -윤해동, 『“만보산 사건”과 동아시아 ‘기억의 터’ -한국인들의 기억을 중심으로-』, 『사이(SAI)』 제14호, 국제한국문학문화학회, 2013, 504면.

예컨대 ‘작가의 심각한 내적 변모와 모색의 결과가 아니라, 만주경영이라는 제국주의의 새로운 시대적 흐름에 편승한, 다시 말해 당대의 국책(國策)에 적극적으로 부응한 소설’¹¹⁾로 평가받는 한편 ‘자의반타의반 조선을 떠나 국책이민으로 이주한 만주에서 부딪치는 피식민자의 참담한 현실 속에서 재만조선인의 생존 문제를 통해 민족협화, 선만일여, 내선일체의 식민주의 담론이 갖는 허구성을 문제 삼고 있는’¹²⁾ 작품으로 여겨지기도 하였다. 최근에는 ‘만주와 그 정부, 그리고 원주민들을 개척의 대상으로 보는 식민주의적 시각으로부터 거리를 확보하지 못한’¹³⁾ 작품이라는 평가를 받기도 하였다. 다수의 문학 연구자들이 <농군>의 문학적 가치와 의미에 관한 비평을 제기해왔으며, 그 과정에서 첨예한 토론 또한 이루어져 왔음을 확인할 수 있다.

만보산 사건을 다룬 한국의 대표적 소설이 이태준의 <농군>이라면, 리훤잉(李輝英)의 <만보산(萬寶山)>은 중국 작가의 대표적 작품이다. 이태준의 <농군>만큼 많은 연구가 한국에서 이루어진 것은 아니지만, 이 작품에 대한 연구자들의 관심은 꾸준히 지속되고 있는 상황이다. 한중일 세 나라의 작품을 대상으로 한 비교 연구도 있고¹⁴⁾, 리훤잉(李輝英)의 <만보산>을 독립적으로 연구한 성과도 있다¹⁵⁾. 기존 연구에 의하면, 리

11) 김철, 「몰락하는 신생(新生) -‘만주’의 꿈과 『농군』의 오독(誤讀)」, 『상허학보』 제9권, 상허학회, 2002, 124면.

12) 고명철, 「동아시아 반식민주의 저항으로서 일제말의 ‘만주 서사’ -이태준의 <농군>과 한설야의 <대륙>을 중심으로-」, 『한국문학논총』 제49집, 2008, 한국문학회, 91면.

13) 김준현, 「한국의 문학지식 장에서 ‘만보산 사건’이 기억되어 온 몇 가지 방식」, 『한국문학연구』 제51호, 동국대학교 한국문학연구소, 2016, 51면.

14) 대표적인 논문으로, 김창호, 「동아시아 ‘타자’ 형상 비교 연구- 만보산 사건을 수용한 한중일 소설을 중심으로」, 『중국현대문학』 제31호, 한국중국현대문학학회, 2004, 379~404면’을 거론할 수 있다.

15) 최근 발표된 논문으로, ‘장영우, 「민족 갈등에서 계급 투쟁으로의 전화(轉化) -이휘영, 『만보산(萬寶山)』론-」, 『한국문학연구』 제51호, 동국대학교 한국문학연구소, 2016, 107~140면’을 거론할 수 있다.

휘영(李輝英)의 <만보산>은 만보산 사건의 전말을 가급적 정확하게 알고자 한 의도에서 창작된 작품이다. 만보산 사건과 관련된 여러 인물의 실제 이름을 그대로 사용하였으며, 작품에 묘사된 사건 전개 또한 실제 상황과 크게 어긋나지 않는다. 조선에서 일어난 배화-학살 사건까지 내용에 포함하고 있으며, 당시 일본의 대륙침략 의도를 직접적으로 고발하는 서술을 전개한 것도 이태준의 <농군>이나 이토 에이노스케의 <만보산>에는 없는 이 작품만의 특징이라 할 수 있다.¹⁶⁾

이는 이태준, 안수길, 장혁주 등이 식민 당국의 검열을 고려하면서 창작을 할 수밖에 없었던 데 반해, 리휘영(李輝英)은 사건의 실체에 가장 가까운 작품을 쓸 수 있는 위치에 있었기 때문이다. 한편 리휘영(李輝英)의 <만보산>이 온전히 객관적 사실만을 내포하고 있는 것은 아니다. 중국과 일본경찰의 폭력적 행위 등을 소문으로 처리하거나 전지적 서술자의 요약적 진술과 개입이 지나치게 많고, 역사 현실의 갈등을 계급의식으로 전화하기 위해 불가피하게 역사 사실에 충실하지 않은 장면으로 소설을 끝내는 것은, 이 작품이 리얼리즘 소설로서 안고 있는 취약점이기도 하다.¹⁷⁾

이태준의 <농군>과 리휘영(李輝英)의 <만보산>은 작가가 처한 환경과 현실 인식 그리고 창작된 시기가 다름에도, ‘만보산 사건’이라는 동일한 역사적 사건을 소재로 서술된 작품이기에, 작품의 서사에서 유사한 정보 요소가 빈번하게 등장한다. 이와 관련해 그동안 <농군>과 <만보산>을 대상으로 한 연구가 꾸준히 이루어져 온 것은 맞지만, ‘만보산 사건’이나 ‘만주의 현실’ 그리고 ‘민족의식’에 초점을 두고 두 작품을 들여다 본 경우가 대부분이지, 소설로서 두 작품의 서사적 특질이나 내재적 정보를

16) 장영우, 『민족 갈등에서 계급 투쟁으로의 전화(轉化) -이휘영, 『만보산(萬寶山)』론-』, 『한국문학연구』 제51호, 동국대학교 한국문학연구소, 2016, 121면.

17) 위의 글, 134면.

본격적으로 비교해 그 의미를 해석한 연구는 없었다. 또한 두 작품의 원문 텍스트를 대상으로, 디지털 환경에서 데이터를 구축·분석·시각화한 연구 사례도 없는 상황이다.

본 연구에서 시도하고자 하는 디지털 인문학 연구 방법론을 활용해 두 작품의 내용을 정밀하게 비교·분석하고 해당 데이터를 공유한다면, 추후 이태준의 <농군>과 리휘영(李輝英)의 <만보산>을 연구하고자 하는 연구자들에게 직접적인 도움을 줄 수 있을 것이다. 뿐만 아니라, 디지털 방법론을 활용해 근현대 소설을 탐구하고자 하는 연구자들에게도 데이터를 매개로 소설에 접근하는 새로운 시각을 제공할 수 있을 것이다.

3. 연구방법: 시맨틱 데이터 편찬 및 그래프 시각화

1) 시맨틱 데이터 편찬 방법론의 활용

시맨틱 데이터(semantic data)는 기본적으로 RDF, RDFS, OWL로 대표되는 시맨틱 웹 언어로 기술된 Triple(S-V-O) 형식의 데이터를 가리킨다. 온전한 형식으로 구축된 시맨틱 데이터는 곧 링크드 데이터(linked data)로 연장되며, 데이터 처리 과정에서 SPARQL과 같은 언어를 통해 의미적 정보가 담긴 질의응답을 활용할 수 있다. 이와 같은 기술적 처리는 기계가독형(machine-readable) 데이터 구축을 바탕으로 하는 것이며, 문서의 웹이 아니라 데이터의 웹으로서 소위 시맨틱 웹(Semantic Web)의 철학을 지향하는 것이자 동시에, 결과물로서 지식그래프(knowledge graph) 구현을 전제하는 것이기도 하다.¹⁸⁾

18) 시맨틱 데이터 편찬 방법론에 관한 설명으로, 장문석·류인태, 『디지털 인문학과 한국 문학 연구(1) : 작가 연구를 위한 시맨틱 데이터베이스 설계』, 『민족문학사연구』 제75권, 민족문학사연구소, 2021, 360면.' 내용을 참고할 수 있다.

이러한 접근은 웹 환경 또는 데이터 처리 과정에 ‘의미(semantics)’를 부여하고자 하는 노력과 맞닿아 있다. ‘의미’의 부여는 소위 빅데이터(big data)로 언급되는 양적(quantitative) 차원의 데이터 처리에서 벗어나, 질적(qualitative) 차원에서 데이터에 접근할 수 있는 채널을 열려는 시도와 밀접한 관련이 있다. 인문학 연구에 이러한 시각을 접목할 경우, 질적 연구에 있어 데이터 처리 과정을 활용하는 것이 가능해진다.¹⁹⁾ 예컨대 ‘멀리서 읽기(distant reading)’로 표방되는 데이터 기반의 문학 독해 방법론은 인간 연구자가 처리하지 못했던 대규모 문학 자료를 데이터로 구축해 계량적 관점에서 분석하는 데 초점을 두지만, ‘의미’ 부여를 지향하는 시맨틱 데이터 프로세싱 방법론은 규모보다는 오히려 질적 차원의 문제의식에 더 주목함으로써 스몰데이터(small data) 기반의 ‘꼼꼼히 읽기(micro reading)’를 지향한다. 규모가 크지 않은 자료를 대상으로 한다 하더라도, 연구자의 문제의식이 분명하게 반영된 데이터 모델링 과정을 거쳐 확보한 시맨틱 데이터는, 논증에 있어 객관성을 도모하고 지식을 공유하는 방식에 있어서도 효과적일 수 있기 때문이다.

본 연구는, SPARQL Endpoint와 같이 시맨틱 데이터를 직접 공유할 수 있는 환경을 마련하는 데 초점을 두기보다는 지식그래프를 구현하고 그것을 분석함으로써 데이터 사이의 의미적 관계를 변별하는 데 집중하고자 하였다. 이와 같은 이유로 시맨틱 데이터 편찬을 지향하되 RDF나

19) 디지털 인문학 연구 방법론으로서 시맨틱 데이터 프로세싱을 인문학 연구에 본격적으로 활용한 국내의 대표적 연구 사례로, ‘지암일기 디지털 인문학 연구’를 거론할 수 있다. 해당 연구는 17세기 양반 윤이후의 『지암일기』를 대상으로 약 8만2천여 건 규모의 시맨틱 데이터를 편찬해서 여러 가지 각도에서 해당 데이터를 시각화·해석한 사례이다. 해당 연구 결과물은 여러 형식의 시맨틱 데이터베이스와 시각화 구현 결과물을 종합한 웹사이트(<http://jiamdiary.info/>)로 제공되고 있으며, 연구 방법 및 과정과 결과물을 정리한 논문(‘류인태, 『데이터로 읽는 17세기 재지사족의 일상 -『지암일기(1692-1699)』 데이터베이스 편찬 연구-』, 한국학중앙연구원 한국학대학원 박사학위논문, 2019.)도 발표되었다.

OWL과 같은 시맨틱 웹 기반의 데이터를 직접적으로 구축하지 않고, Triple(S-V-O) 형식의 결과물 데이터셋을 취합과 동시에 그것을 지식그래프로 표현하기에 용이한 환경을 활용하는 데 집중하였다. 이에 따라 해당 조건에 최적화된 데이터 처리 플랫폼으로서 그래프데이터베이스 Neo4j²⁰⁾를 연구 과정에 적용하였다.

2) 시맨틱 데이터 모델링: 클래스와 속성 및 관계의 정의

시맨틱 데이터는 Triple(S-V-O) 형식을 갖추기 위해 기본적으로 온톨로지(Ontology)²¹⁾ 설계를 전제로 구축된다. 온톨로지의 구성 요소 가운데 클래스(class), 속성(property), 관계(relation)에 대한 정의는 필수적이며, 이에 따라 본 연구는 연구대상으로서 이태준의 <농군>과 리혜영(李輝英)의 <만보산>을 비교하기 위한 주요한 문제의식을 바탕으로 클래스와 속성 그리고 관계를 다음과 같이 정의하였다.

20) Neo4j(<http://www.neo4j.com>)는 Neo Technology사에서 개발한 그래프 데이터베이스 플랫폼이다. Cypher Language를 질의어로 삼고 있으며, Triple(S-V-O) 형식의 시맨틱 데이터를 그래프 형태로 출력하는 데 용이하기에, 인문학 자료를 대상으로 한 시맨틱 데이터베이스 구축에 있어서 여러모로 높은 활용도를 지닌 도구에 해당한다. -Neo4j에 대한 설명은 ‘장문석·류인태, 『디지털 인문학과 한국문학 연구(1) : 작가 연구를 위한 시맨틱 데이터베이스 설계』, 『민족문화사연구』 제75권, 민족문화사연구소, 2021, 411면’ 각주 내용을 참고하였다.

21) 온톨로지(ontology)는 현실 세계에서 관찰되는 대상 사이의 관계 및 여러 속성을 컴퓨터가 이해하고 처리할 수 있도록 표현한 체계로서, 클래스(class), 개체(instance), 속성(property), 관계(relation) 4가지 구성 요소로 표현된다. 온톨로지를 작성하는 대표적인 언어로, 웹상의 정보 처리에 활용되는 웹 온톨로지 언어(OWL, Web Ontology Language)를 거론할 수 있다. 온톨로지는 인공지능(AI), 시맨틱 웹(Semantic Web), 자연어처리(NLP) 등 다양한 분야에서 지식을 처리하고 공유하기 위한 맥락으로 활용되고 있다. -온톨로지에 대한 설명은 ‘류인태, 『데이터 기반의 고전 읽기 교육 — 『논어』를 대상으로 한 디지털 인문학 강의 사례를 중심으로』, 『인문논총』 제78권1호, 서울대학교 인문학연구원, 2021, 47면’ 각주 내용을 참고하였다.

(1) 클래스(class)와 속성(property) 정의

〈표1〉 클래스와 개체 예시

클래스	설 명	개체 예시
작품 (Work)	연구대상으로서 구체적인 작품	예시1: 농군 예시2: 만보산
플롯 (Plot)	소설 속 서사 전개에의 가장 큰 구성요소	예시1: 조선인들의 만주 도착 예시2: 황무지 임대
사건 (Event)	서사 전개를 위해 배치된 이야기의 최소 단위	예시1: 중국인과의 충돌 예시2: 첫 번째 농민집회
인물 (Figure)	소설에 등장하거나 언급되는 인물	예시1: 윤장권 예시2: 하오용더(郝永德)
장소 (Place)	소설에 언급되거나 구체적 사건이 발생하는 장소	예시1: 봉천행 기차 예시2: 장푸(張福) 집
신분 (Status)	소설에 등장하는 여러 인물의 계급과 직업 등 신분적 요소	예시1: 소작인 예시2: 지주
행위 (Action)	소설에 등장하는 인물이 특정 사건에서 취한 행위	예시1: 때리다 예시2: 뇌물을 주다

온톨로지 상의 클래스는 그래프데이터베이스(Neo4j)에서 라벨(label)에 해당한다. 위의 <표1>에서 확인할 수 있듯이 모두 7가지 클래스를 설계하였으며, 그에 대한 정의 맥락과 개체 속성은 다음과 같다.

① 작품(Work) 클래스는, 연구대상으로서 구체적인 작품을 정의하기 위한 범주이다. 현 시점에서 이 클래스에 속하는 개체는 이태준의 <농군>과 리훤잉(李輝英)의 <만보산> 두 작품밖에 없지만, 추후 연구 범위를 확장해서 여타 작품들 또한 비교를 위한 개체로 삼을 수 있다는 전제하에 설계되었다.

작품 클래스의 개체 속성은 식별자(id), 대표명(name), 작가(author), 간행처(publisher), 발간연도(pYear), 총쪽수(pages)로 구성된다.

② 플롯(Plot) 클래스는, 말 그대로 두 소설의 이야기 전개 과정에 있어서 기능하는 가장 큰 구성 요소를 정의하기 위한 범주이다. 소설 속 이야기를 분절하는 가장 큰 단서이자, 개별 작품을 구성하는 가장 큰 개체에 해당한다. 작품의 서사에 나타나는 기승전결의 흐름에 의거할 때 거시적 차원에서 안배할 수 있는 소설의 가장 큰 구성단위이기에, 본 클래스에 해당하는 개체 정보의 구축은 매우 조심스럽고 중요한 작업이다.

플롯 클래스의 개체 속성은 식별자(id), 대표명(name), 쪽수(page)로 구성된다.

③ 사건(Event) 클래스는, 소설에서 전개되는 서사 구성의 최소 단위를 정의하기 위한 범주이다. 앞서 플롯(Plot) 클래스가 작품 전체의 서사 전개를 고려한 가장 큰 이야기 단위라면, 사건 클래스는 개별 플롯을 구성하는 가장 작은 단위의 이야기에 해당한다. 플롯과 사건 클래스는 소설의 전체 서사 흐름과 그것을 뒷받침하는 구체적인 갈등 또는 이야기를 데이터로 분절한 것으로서, 두 클래스에 소속되는 개체의 관계 양상을 통해 작품의 전체적 서사 구성 및 그 특징을 확인할 수 있다.

사건 클래스의 개체 속성은 식별자(id), 대표명(name), 설명(description), 분류(sort)로 구성된다. 이 가운데 분류는 개별 사건의 성격을 지정하기 위한 속성으로서, 갈등에 해당하는 사건의 경우 ‘갈등’, 갈등이 해소되는 맥락의 사건은 ‘갈등 해소’ 그 외의 사건은 모두 ‘일반’으로 속성값(value)을 고정하였다.

④ 인물(Figure) 클래스는, 소설에 등장하거나 언급되는 인물을 정의하기 위한 범주이다. 사건 클래스를 기준으로 볼 경우, 구체적 사건을 주동하거나 보조하는 주체이기도 하다.

인물 클래스의 개체 속성은 식별자(id), 대표명(name), 한자명(chiname),

국적(nation), 유형(type)으로 구성된다. 이 가운데 유형은 개별 인물의 유형을 구분하기 위한 속성으로서, 고유명사로 지칭되는 특정한 인물의 경우 ‘특정인’, 고유명사가 아니라 일반명사로 지칭되는 인물(군)의 경우 ‘일반인’으로 속성값(value)을 고정하였다.

⑤ 장소(Place) 클래스는, 소설에 언급되거나 구체적인 사건이 일어나는 장소를 정의하기 위한 범주이다. 본 연구에서 구체적 대상으로 삼은 두 작품은 ‘만보산 사건’을 공통적 소재로 삼았으나, 해당 사건이 어떻게 전개되어 나갔는지 그 양상을 구체적으로 묘사할 수 있는 매개로서 장소가 다르게 설정되었다는 사실은, 두 소설을 비교하는 데 있어서 중요한 단서가 된다.

장소 클래스의 개체 속성은 식별자(id), 대표명(name)으로 구성된다.

⑥ 신분(Status) 클래스는, 소설에 등장하는 여러 인물의 계급과 직업 등 신분적 요소를 정의하기 위한 범주이다. 두 작품의 서사에서 주요한 역할을 하는 인물들의 신분은 곧 실제 ‘만보산 사건’을 일으키고 전개한 주체가 누구냐의 문제와 관련해 매우 중요한 정보가 되며, 이는 작가의 현실 인식과도 연결되기에 작품을 해석하는 데 있어서도 핵심적 단서가 될 수 있다.

신분 클래스의 개체 속성은 식별자(id), 대표명(name)으로 구성된다.

⑦ 행위(Action) 클래스는, 소설에 등장하는 인물이 특정 사건에서 취한 행위를 정보화하기 위한 목적에서 설계된 범주이다. 작품 속에 묘사된 등장인물의 행위는 인물들의 ‘신분(Status)’ 만큼이나 작품 속 사건과 서사 전개의 연결 고리를 이해하는 데 있어서 중요한 정보이다. 특정 사건에 연루된 특정 인물이 어떠한 행위를 했고, 그 행위가 해당 사건을 어떠

한 방향으로 흘러가게 하느냐는 곧 작품 전체 서사의 줄기로서, 작가의 시각을 간접적으로 확인할 수 있는 유효한 단서가 되기 때문이다.

행위 클래스의 개체 속성은 식별자(id), 대표명(name)으로 구성된다.

7가지 클래스 가운데 ①작품(Work) ②플롯(Plot) ③사건(Event) 클래스는 두 소설의 서사 전반을 데이터로 재구성하기 위한 목적으로 설계된 것이며, ④인물(Figure) ⑤공간(Place) ⑥신분(Status) ⑦행위(Act) 클래스는 서사에 개입하는 주요한 문맥 정보(contextual information)가 어떠한 변별성을 지니면서, 그것들 사이의 의미적 관계가 서사에 미치는 영향이 무엇인지를 구체적으로 살펴보기 위해 설계된 것이다.

(2) 관계(relationship) 정의

7가지 클래스에 해당하는 개체 정보는 지식그래프 상에서 노드(node)를 구성한다. 각각의 노드를 연결하는 링크(link)를 구현하기 위해서는 온톨로지 상에서 관계(relationship)를 설계해야 한다. 앞서 서술하였듯이 본 연구에서 클래스를 설계하는 가운데 초점을 둔 것은, 소설 속 서사 전개를 어떻게 섬세하게 정보화할 수 있느냐의 문제와 서사를 구성하는 의미적 요소로서 다양한 문맥 정보를 어떻게 효과적으로 연결할 수 있느냐의 문제이다. 이러한 문제의식에 입각해 아래와 같은 10가지 의미 관계를 정의하였다.

〈표2〉 관계 정의

관계	설명	Domain(Class, A)	Range(Class, B)
isPartOf	A가 B에 포함되다	Plot	Work
		Event	Plot
isRelatedTo	A가 B와 연관되다	Event	Event
takesPlaceIn	A가 B에서 발생하다	Event	Place

appearsIn	A가 B에 등장하다	Figure	Event
hasParents	A는 B의 자식이다	Figure	Figure
hasSpouse	A는 B의 배우자다		
hasAcquaintance	A는 B의 지인이다		
instructs	A가 B에게 지시하다		
appearsAs	A가 B로 등장하다	Figure	Status
acts	A가 B를 실행하다	Figure	Action

① ‘A가 B에 포함되다(isPartOf)’는 작품(Work)과 플롯(Plot) 그리고 플롯(Plot)과 사건(Event) 사이의 위상을 정의하기 위한 관계이다. 예컨대 특정 작품을 구성하는 개별 플롯과 각각의 플롯을 구성하는 개별 사건 사이의 위상 관계를 가리킨다.

예시: 황무지 임대(Plot) isPartOf 만보산(Work), 매국노라는 고민(Event) isPartOf 황무지 임대(Plot)

② ‘A가 B와 연관되다(isRelatedTo)’는 사건(Event)과 사건(Event) 사이의 인과적 개연성을 정의하기 위한 관계이다. 사건 사이의 인과성을 데이터로 정리하는 목적은 곧 플롯을 통해 구성되는 서사적 특질을 객관화하기 위한 시도와 관련이 있다.

예시: 중국인과의 충돌(Event) isRelatedTo 공사 시작(Event)

③ ‘A가 B에서 발생하다(takesPlaceIn)’는 사건(Event)과 해당 사건이 일어난 장소(Place)를 의미적으로 연결하기 위한 관계이다. 두 작품에 각기 묘사된 사건과 장소의 변별 및 두 요소의 관계적 양상은, 작품의 특징을 비교하기 위한 중요한 단서가 된다.

예시: 첫 번째 농민집회(Event) takesPlaceIn 마바오산(馬寶山)의 집(Place)

④ ‘A가 B에 등장하다(appearsIn)’는 인물(Figure)과 해당 인물이 등장한 사건(Event)을 의미적으로 연결하기 위한 관계이다. 서사를 이끌어 나가는 주체로서 등장인물이 작품 속 갈등에 미치는 영향의 양상은, 곧 개별 사건의 전개 및 사건들 사이의 인과적 양태와 직결된다. 이러한 지점은 소설을 이해하는 데 있어서 중요한 단서가 된다.

예시: 창권(Figure) appearsIn 기차 안의 대화(Event)

⑤ ‘A는 B의 자식이다(hasParents)’, ‘A는 B의 배우자다(hasSpouse)’, ‘A는 B의 지인이다(hasAcquaintance)’, ‘A가 B에게 지시를 내리다(instructs)’ 등은 인물(Figure)과 인물(Figure) 사이를 정의하기 위한 관계이다. 소설 속에 등장하는 인물들 사이의 관계는 대체로 혈연, 혼연, 지연, 학연 및 직장 동료나 애인에 이르기까지 다양한 연결고리를 보여준다. 본 연구에서 정의한 관계는 이태준의 <농군>과 리훤잉(李輝英)의 <만보산> 두 소설 속 등장인물들 사이에 나타나는 관계로서, 작품 속 갈등 및 권력 관계를 파악하기 위한 기본적 단서가 된다.

예시: 김복(Figure) hasParents 김복 아버지(Figure), 창권(Figure) hasSpouse 창권 아내(Figure), 나카가와(Figure) instructs 하오용더(郝永德)(Figure)

⑥ ‘A가 B로 등장하다(appearsAs)’는 인물(Figure)과 해당 인물의 계급이나 직업과 같은 신분(Status)을 의미적으로 연결하기 위한 관계이다. 인물의 신분 정보는 두 소설의 시대적 배경을 이해하는 데 기본 정보가 되며, 중심 갈등인 ‘만보산 사건’의 전개에 영향을 미치는 인물들의 사회적 위상을 파악하는 데 있어서 또한 중요한 단서가 된다.

예시: 창권(Figure) appearsAs 소작인(Status)

⑦ ‘A가 B를 실행하다(acts)’는 인물(Figure)과 해당 인물의 특정 행위(Act)를 의미적으로 연결하기 위한 관계이다. 특정 사건에 연루된 특정 인물의 행위는 해당 사건의 향방에 직접적 영향을 미치는 요소로서, 곧 서사의 흐름과 연계해 작가의 관점을 드러내는 유효한 정보에 해당한다.

이 관계어의 경우 ‘기색(mood)’이라는 관계 속성을 부여하였다. 기색은 개별 행위에 내포된 눈치나 낯새 또는 감정의 맥락을 정의하기 위한 속성으로서, 각각의 행위 이면에 자리한 정서적 상태를 속성값(value)으로 부여하여, 동일한 행위라 하더라도 다른 의미 맥락을 정의하고자 하였다.

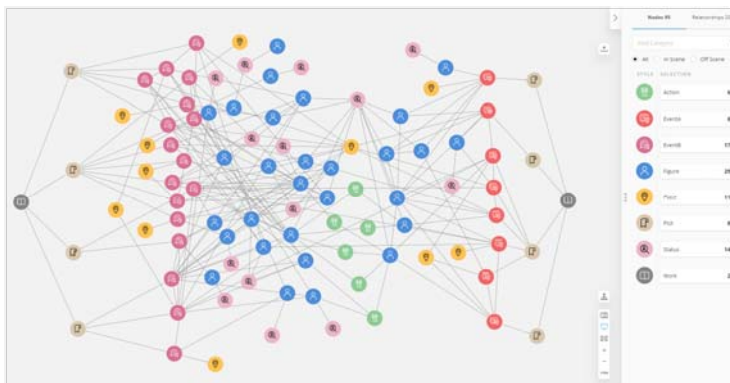
예시: 황채심(Figure) acts(mood:“기대”) 연설(Act)

3) 편찬된 시맨틱 데이터의 규모와 형태

〈표3〉 시맨틱 데이터의 전체 규모

클래스 A	노드 데이터(A)	클래스 B	링크 데이터(A→B)
작품(Work)	2	·	
플롯(Plot)	8	작품(Work)	8
사건(Event)	25	플롯(Plot)	25
		사건(Event)	18
		장소(Place)	25
인물(Figure)	29	사건(Event)	81
		인물(Figure)	8
		신분(Status)	29
		행위(Action)	13
장소(Place)	11	·	
신분(Status)	14	·	
행위(Action)	6	·	
노드 데이터 총 95 건		링크 데이터 총 207 여건	
총 302 건			

앞서 정리한 클래스(class), 속성(property), 관계(relation)의 설계 내용을 바탕으로 이태준의 <농군>과 리훤잉(李輝英)의 <만보산>에서 데이터를 추출·구축하였다. 전체 데이터 규모는 <표3>과 같다.²²⁾ 총 300여 건의 데이터는 그래프데이터베이스(Neo4j)를 활용해 지식그래프 형태로 구현하였으며, 아래 [그림1]은 해당 지식그래프의 전체 외형이다.



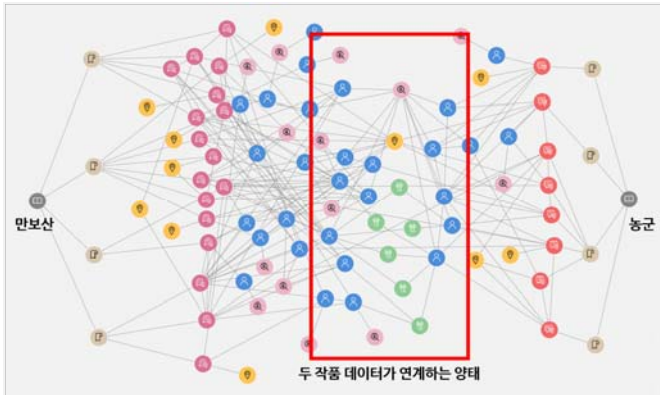
[그림1] 전체 데이터를 네트워크 그래프로 출력한 모습: Neo4j Bloom

4. 연구결과: 지식그래프를 활용한 작품 비교

<농군>과 <만보산> 두 작품에서 추출한 300여 건의 데이터를 지식그래프로 구현한 결과물 [그림2]을 볼 경우, 좌측에 위치한 <만보산>의 플롯, 사건, 인물 데이터가 우측에 위치한 <농군>의 플롯, 사건, 인물 데이터에 비해 상대적으로 그 양이 많고 관계 양상 또한 복잡하다는 것을 확인할 수 있다. 두 작품의 전체 분량이나 서사의 복잡함 그리고 등장인물

22) 구체적인 데이터셋은 본 연구의 온라인 페이지(<http://www.klbksk.com/wiki/index.php/NovelsAboutManbosanIncident>)의 'Dataset' 항목을 통해 확인할 수 있다.

의 숫자가 다른 것이 편찬된 데이터를 통해서 드러나는 것이다. 작품(Work)이 플롯(PLOT)으로 플롯이 사건(Event)으로 다시 사건이 인물(Figure)과 장소(Place)로 연결되는 개체 데이터의 의미적 관계를 고려할 시, 두 작품을 중심으로 한 데이터가 서로 만나는 지점은 [그림2]의 빨간 사각형이 그려진 곳, 즉 그래프의 중앙 즈음이라는 것을 알 수 있다.



[그림2] 「농군」과 「만보산」 데이터가 그래프 상에서 연계하는 양상

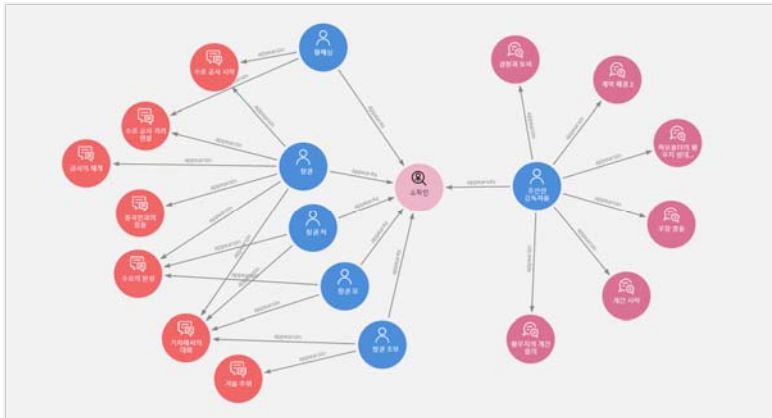
빨간 사각형으로 표시된 부분은 두 작품을 통틀어 가장 많은 관계로 연결된 인물(Figure), 행위(Action), 신분(Status), 장소(Place) 등이 위치해 있다. <농군>과 <만보산>의 비교가 본 연구의 목적이기에 지식그래프를 탐색하는 맥락은 해당 노드들에 초점을 두었으며, 그 가운데에서도 신분 개체 '소작인'과 행위 개체들이 상대적으로 두드러진 노드 데이터로 확인되었다.

1) 두 작품에서 공통적으로 다루는 신분(Status): 소작인

<농군>과 <만보산>의 데이터를 비교하면 인물(Figure)과 신분

(Status) 데이터가 연계하는 양상에서 차이점이 가장 뚜렷하다. 예컨대 두 작품 모두 조선인 농부가 등장하는데, 조선인 농부가 그려지는 양태는 세부적으로 조선 소작인과 조선 쿨리 그리고 노동자로 분류할 수 있다. 이 가운데 두 작품의 가장 큰 차이점이 나타나는데, 조선 소작인이라는 신분이 작품의 서사에서 맡고 있는 역할이 바로 그것이다. 조선 소작인을 매개로 데이터를 출력해보면 아래 [그림3]과 같은 결과를 확인할 수 있다.

```
Cypher Query: match (a:Status{name:'소작인'}) <- [r] - (b{nation:'조선'}) - [l] ->
(c{class:'Event'}) return a, r, b, l, c
```



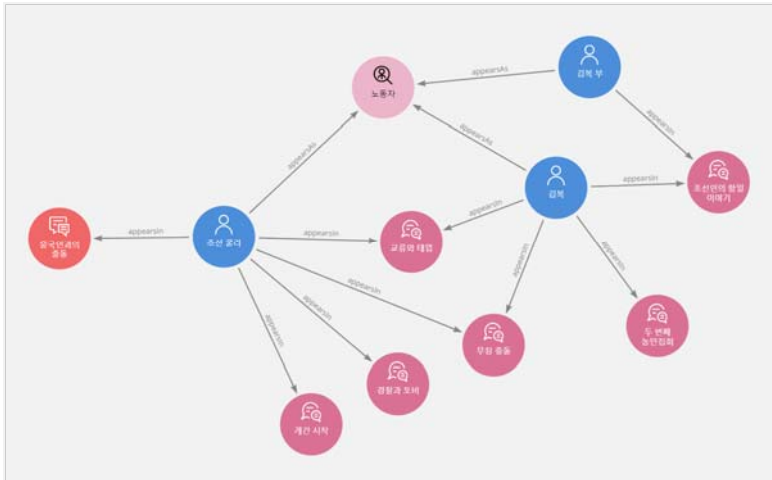
(그림3) 두 작품에 등장하는 조선 소작인과 연결된 사건을 출력한 결과

<농군>의 주인공 창권은 만주 지역으로 이주한 조선 소작인인데, 해당 신분은 소설의 중심 화자로서 창권의 역할을 대변한다. <만보산>의 경우, 중국 소작인의 시선에 입각해 조선 소작인과 조선 노동자를 응시하는데, 전자는 부정적으로 바라보는 반면, 후자는 긍정적으로 바라본다. 이에 대해 구체적으로 살펴보면, <농군>에 묘사된 조선 소작인은 만보산 사건을 일으킨 중심 원인인 수로 공사와 그로부터 야기된 중국 토착민에

대한 대응 그리고 무장 충돌에 이르기까지 그 모든 사건의 주체로 그려진다. 그에 비해 <만보산>에 묘사된 조선 소작인은 수로 공사에 직접 참여하지 않고 감독자의 신분에 가까우며, 연결된 사건의 양상을 살펴보면 실제 작업을 하는 조선인 노동자들을 감시하고 심지어 학대까지도 일삼는다. 신분상으로는 소작인으로 등장하지만 실제 행위는 농사를 짓는 사람이 아니라 다른 노동자를 부리는 지주 이미지에 더욱 가까워 보인다.

한편 <만보산>에 묘사된 조선 노동자는 <농군>에서 주인공 창권이 속한 조선 소작인의 이미지와 가깝게 보인다. 이와 관련해 조선 노동자에 대한 묘사가 <농군>에 거의 없는 것도 주의할 만한 사실이다. 데이터를 다뤄보면 더 많은 사실을 확인할 수 있다.

Cypher Query: match (a:Status{name:'노동자'}) <- [r] - (b{nation:'조선'}) - [] -> (c{class:'Event'}) return a, r, b, l, c



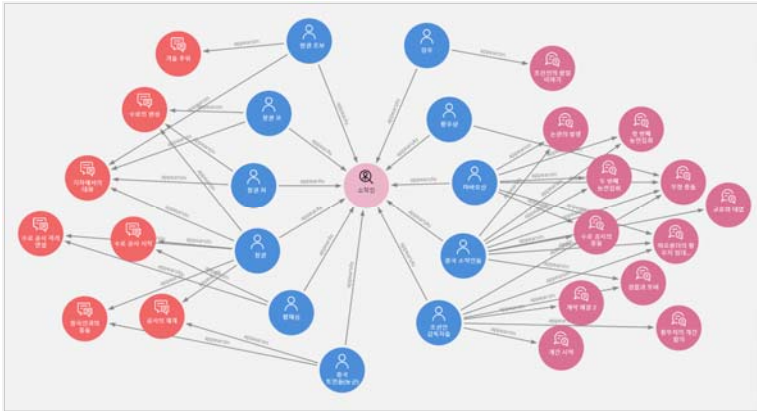
(그림4) 두 작품에 등장하는 조선 노동자와 연결된 사건을 출력한 결과

두 작품에서 조선 소작인과 노동자를 바라보는 화자의 시선은 매우 다르다. <만보산>의 화자는 조선 소작인을 적으로 상정하는 한편 조선 노동자의 경우 항일 활동을 함께 전개할만한 전우로 상정한다. 이와 관련해 작품 속에는 중국 농민들이 황무지를 개간하는 조선인을 바라보며 ‘이들은 나쁜 조선인인가 좋은 조선인인가’라는 뉘앙스로 논의하는 장면도 나오는데, 이를 통해 당시 중국인들은 조선의 이주민을 일괄적으로 바라보지 않았으며, 행위의 결과에 따라 좋은 사람과 나쁜 사람으로 구분해 판단하였음을 짐작할 수 있다. 이와 같은 시각은 황무지 개간에 필요한 수로 공사 문제로 인해 발생한 조선 농민과 중국 농민 사이의 갈등을, 민족 문제가 아니라 계급의식의 맥락으로 전환함으로써 반제국주의·항일의 차원에까지 해당 문제를 연장한 것이라 할 수 있다.

이를 위해 리훤잉(李輝英)은 작품에서 중국 학생 리징평(李竟平)을 완벽한 인물로 그려내고자 하였으며 조선 노동자 김복의 경우 일제로부터 도망하며 적극적으로 저항하는 인물로 묘사함으로써, 개간사업에 참여한 조선노동자들을 중국농민과 연합하게 하는 서사 전략을 선택한다.²³⁾ 상대적으로 조선 노동자에 대한 묘사가 <농군>에 거의 없을 정도로 부족한 것은, 소작인과 노동자 모두 수로 건설에 동참하는 단결된 조선인으로 그려진 탓이다. 리훤잉(李輝英)의 <만보산>이 계급주의적 시각에 상대적으로 기울어져 있다면, 이태준의 <농군>은 민족주의적 시각이 우위에 있음을 짐작할 수 있다.

23) 장영우, 「민족 갈등에서 계급 투쟁으로의 전화(轉化) -이휘영, 『만보산(萬寶山)』론-」, 『한국문학연구』 제51호, 동국대학교 한국문학연구소, 2016, 108면.


```
Cypher Query: match (a:Status{name:'소작인'}) <- [r] - (b) - [] -> (c(class:'Event'))
return a, r, b, l, c
```



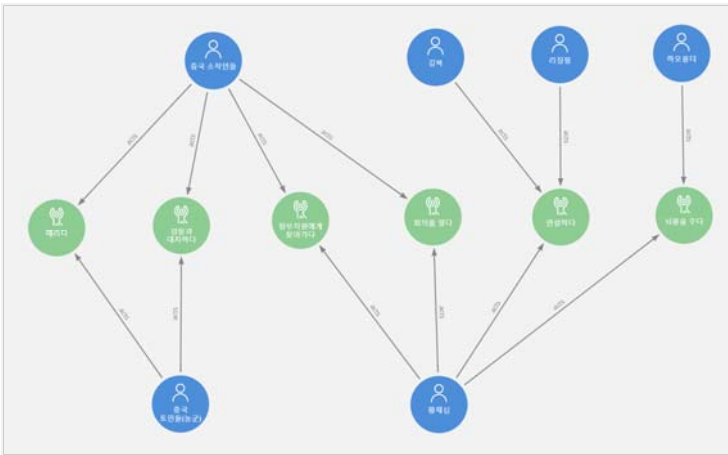
[그림6] 두 작품에 등장하는 소작인과 연결된 사건을 출력한 결과

[그림6]은 두 작품 모두에서 소작인에 해당하는 인물들과 그들이 등장하는 사건을 출력한 결과이다. 좌측은 <농군>의 인물과 사건, 우측은 <만보산>의 인물과 사건들인데, 두 작품 모두 각기 조선과 중국의 소작인을 사건 진행에 있어서 중요 인물로 상정하였음을 확인할 수 있다.

2) 두 작품에서 공통적으로 나타나는 ‘행위(Action)’

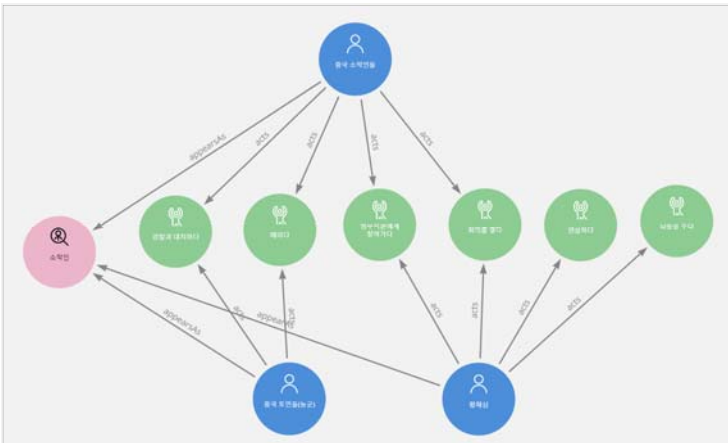
신분(Status) 클래스에서 ‘소작인’ 데이터가 두 작품을 비교하는 데 유효한 역할을 했다면, 행위(Action) 클래스의 경우 개체 데이터 전반이 두 작품 사이에 위치해 있으면서 유의미한 시사점을 던져준다. 이는 ‘만보산 사건’이 두 작품의 서사에 공통적으로 활용되는 과정에서 해당 사건에 참여하는 인물들의 물리적인 행위 양태가 크게 다르지 않은 것으로 묘사되며, 이로 인해 인물들의 행위 이면에 자리한 의도와 그에 따른 서사적 효과를 짐작해볼 수 있기 때문이다.

Cypher Query: `match (a:Figure) - [r:acts] -> (b:Action) return a, r, b`



(그림7) 두 작품에 등장하는 인물들의 행위를 출력한 결과

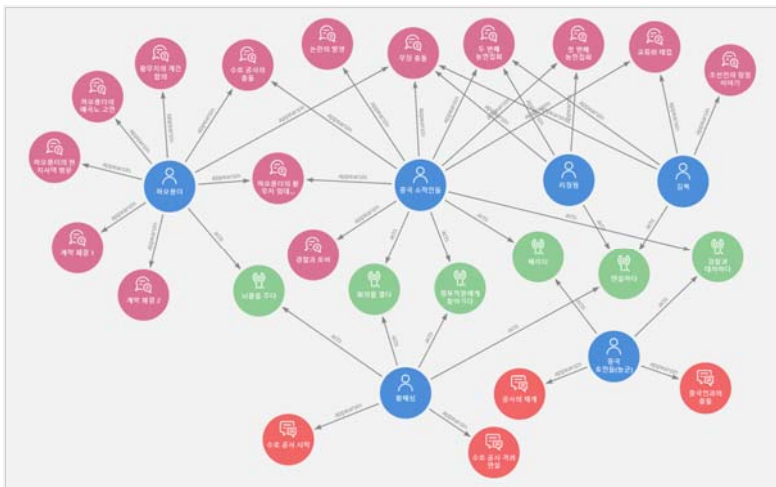
```
Cypher Query: match (a:Status{name:'소작인'}) <- [r] - (b:Figure) - [l] -> (c:Action)
return a, r, b, l, c
```



(그림8) 두 작품에 등장하는 ‘소작인’ 해당 인물들의 행위를 출력한 결과

예컨대 [그림7]은 두 작품에 등장하는 인물들이 공통적으로 수행한 행위를 매개로 유관 데이터를 출력한 결과이며, [그림8]은 두 작품에 등장하는 인물들 가운데 ‘소작인’에 해당하는 인물이 수행한 행위를 매개로 유관 데이터를 출력한 결과이다. 두 결과 모두 가운데 배치된 행위 데이터를 기준으로 상단은 <만보산>의 인물들이며, 하단은 <농군>의 인물들이다. [그림7]을 기준으로 [그림8]을 비교해보면, 동일한 조건에서 ‘소작인’이라는 신분(Status) 조건만을 변수로 부여했을 뿐인데 <농군>은 출력 결과에 차이가 없는 반면, <만보산>은 출력 결과에서 리징핑(李竟平), ‘김복’, ‘하오용더(郝永德)’ 등의 인물이 사라진 것을 확인할 수 있다. 리징핑(李竟平)은 신분이 ‘학생’이며, 김복은 ‘노동자’ 그리고 하오용더(郝永德)는 ‘경영자’이다.

Cypher Query: match (c(class:'Event')) <- [s] - (a:Figure) - [r:acts] -> (b:Action)
return a, r, b, s, c



[그림9] 두 작품에 등장하는 인물들의 행위와 등장 사건을 출력한 결과

이들 3인은 非소작인 신분으로서 ‘뇌물’과 ‘연설’에 관한 행위를 하였는데, 해당 행위들은 양적으로나 질적으로나 중국 소작인들의 행위와 비교할 때 상대적으로 그 의미가 크지 않다고 할 수 있다. 예컨대 양적 관점에서 볼 경우 중국 소작인들의 행위는 모두 4가지이며, 非소작인 3인의 행위는 합쳐서 2가지에 불과하다. 질적 관점에서 짚어 보아도 리징평(李竟平)과 김복의 ‘연설’ 행위는 서사 전개에 있어 그 나름의 기능이 있다고 할 수 있겠으나, 하오용더(郝永德)의 ‘뇌물’ 행위는 서사를 끌어가는 인물의 주요 행위라고 보기 어렵다. 이와 달리 [그림9]에서 확인할 수 있듯이 중국 소작인의 행위 4가지는 사건 진행에 있어 새로운 갈등을 야기함으로써 서사 전개에 중요한 역할을 한다. 행위의 차원에 있어서도 두 작품의 서사를 이끌어나가는 주체적 신분이 ‘소작인’임을 알 수 있는 것이다.

참고로 행위에 속하는 6개의 구체적 데이터는 그 성격을 기준으로 크게 2가지로 나누어 볼 수 있다. 우선 ‘정부직원을 찾다’, ‘뇌물을 주다’ 행위는 모두 중국의 정부 직원을 대상으로 한 것으로서 <농군>과 <만보산>에서 해당 행위의 주체는 각기 다르지만, 행위의 대상이 중국 정부 직원이라는 사실은 동일하다.

한편 ‘연설을 하거나’ ‘회의를 여는’ 행위는 <농군>에서 황채심이 조선인을 모아 회의를 열고 연설을 하며 <만보산>에서는 리징평(李竟平)과 김복이 중국인을 모아 회의를 열고 연설을 한다. 마찬가지로 ‘경찰과 대치하는’ 행위는 <농군>의 경우 조선인이 주체가 되어 중국경찰에 대치하며, <만보산>은 중국인이 일본경찰에 대치하는 것으로 나타난다. 즉, 각 작품에 등장하는 인물들은 모두 ‘우(友)’로 나온 인물을 단결시키는 동시에 ‘적(敵)’으로 나온 인물에 저항한다는 것을 알 수 있다.

그리고 한 가지 더 짚어볼 만한 사실은, 『만보산』에서 ‘때리다’ 행위의 경우 중국 소작인들이 주체가 되어 조선 소작인과 중국 매국노에게 폭력을 가하는 양상이며, ‘경찰과 대치하다’ 행위는 일본경찰 뿐만 아니라 중

국경찰까지 대치의 대상으로 삼는다는 것이다. 같은 중국 사람도 때릴 수 있고 경우에 따라 중국 경찰과 대치 할 수도 있다는 생각은 곧, <만보산>의 작가 리훤잉(李輝英)이 민족의식에서 벗어나 계급의식을 적극적으로 드러낸 징후라고 할 수 있다.

3) 종합적 결론

지식그래프를 매개로 편찬된 데이터를 다각도로 살펴봄으로써 두 작품을 섬세하게 비교해보았다. <농군>과 <만보산>은 각기 조선과 중국 출신 ‘소작인’ 신분을 중심인물로 삼아 사건을 구성하고 서사를 전개해나간다는 것을 확인하였다. 반면 ‘노동자’의 모습은 두 작품에서 모두 비중 있게 묘사되지 않을 뿐더러 <농군>의 경우 직접적인 언급이 거의 이루어지지 않았다. 신분(Status) 뿐만 아니라 행위(action) 데이터를 매개로 두 작품에 등장하는 인물과 사건의 양상을 살펴보았을 때도 마찬가지로 소작인에 해당하는 인물들의 역할이 중요하게 나타났다.

다만 이러한 양상 가운데 두 작품에서 드러나는 작가 의식의 경우 다소 차이가 있다. <농군>에 묘사된 조선인과 중국인 사이의 관계는 무조건적으로 적대적이지만, <만보산>에 등장하는 중국인들은 조선인의 신분과 행위에 따라 다른 태도를 드러낸다. 이는 작가 리훤잉(李輝英)이 중국 지주와 같은 악역으로서 조선 소작인을 그려내고, 한편으로 조선 노동자들은 계급투쟁과 항일운동에 앞장서는 긍정적 집단으로 묘사함으로써 조선인을 소위 ‘적(敵)’과 ‘우(友)’ 두 부류로 나눈 탓이다. <만보산>에 나타나는 이러한 특징적 시각은 조선인뿐만 아니라 중국인 등장인물에게도 마찬가지로 적용되는데, 중국인 가운데도 경찰, 매국노, 지주 등은 ‘적’으로 그려지며 학생은 ‘우’로 묘사된다. 즉 <농군>의 경우 민족주의적 시각이 강하게 드러나는 반면 상대적으로 <만보산>은 계급주의적 시각을 뚜

렷하게 강조한다는 것이, 데이터를 통해 확인할 수 있는 두 작품의 가장 큰 차이점이라 할 수 있다.

두 작품 사이에 이와 같은 차이점은 왜 발생하는가. 그 원인을 추측해보자면 우선 <만보산>은 작가 리휘잉(李輝英)이 1932년부터 1933년 사이에 쓴 그의 초기 작품으로서, 만보산 사건 직후 발표된 소설이다. 이 작품을 쓸 무렵 리휘잉(李輝英)은 ‘중국좌익작가연맹’이라는 문학 단체에 참여하기도 했는데, 이후 <만보산>은 중국 항일문학의 효시로 평가 받게 된다. 이와 같은 배경은 ‘항일’이라는 목표를 달성하기 위한 사회주의적 계급의식의 표출로서, <만보산>이라는 작품을 바라볼 수밖에 없게끔 만든다.

예컨대 소설의 후반부터 결미에 이르는 내용의 경우 실제와는 다른 방향의 사건 전개가 두드러지는데, 이는 <만보산>이 지향하는 바가 단순히 만보산 사건을 온전히 재현하고자 한 것이 아니라 해당 사건을 소재로 삼아 사회주의 이데올로기와 항일 의식을 강조하고자 하였다는 것을 단적으로 보여주는 예라 할 것이다. 소설의 결말부에서 중국 농민들이 엽총을 가지고 조선 노동자들과 연합해 중국과 일본 경찰에게 공격을 시작하고, 예기치 않게 경찰들을 논밭에서 쫓아내는 데 성공하기까지 하는데, 이 지점에서 독자들은 강한 의문이 들 수밖에 없다. 훈련 받지 못한 농민들이 무장충돌에서 어떻게 경찰을 격퇴할 수 있는가? 후반부에 이르면 <만보산>은 일부 허구적 서사 개입을 통해 작가의 의도를 암묵적으로 강조하고 있다는 것을 알 수 있다.

결말부에 이르러 소설로서의 허구성을 은연중 강조하고 있음에도 불구하고, <만보산>은 <농군>에 비하면 상대적으로 만보산 사건의 사실을 있는 그대로 재현하고자 한 작품으로 평가받는다. 그에 비해 <농군>은 만보산 사건에 관한 사실 재현과는 다소 거리가 먼 작품이다. 이태준이 <농군>을 창작한 시기는 1939년으로서, 작품의 서두에서 이 소설의 시

간접 배경은 그 이전인 장작림 정권 시대임을 강조하면서, 만보산 사건을 소재로 삼았음을 밝힌다. 이러한 측면은 1930년대 말 조선인을 만주로 보내던 ‘국책이민’ 정책에 대한 이태준의 숨겨진 대응으로서, 만주에서 중국인 때문에 온전한 삶을 살지 못하는 재만조선인의 모습을 그려냄으로써 ‘민족협화’, ‘선만일여’, ‘내선일체’ 등 만주국에 대한 당시 일본의 선전과 식민주의 담론의 허구성을 문제 삼고자²⁴⁾ 한 단서로 해석할 수도 있다.

예컨대 중국인과 조선인이 만주에서 같이 살 수 없을 정도로 민족적 갈등에 대한 극단적 묘사가 <농군>에서 이루어진 배경에는, 그와 같은 작가의 의도가 있다고 이야기할 수 있는 것이다. 일본에 식민통치를 받고 있던 조선인은 만주 땅에서 중국 토착민들로부터 식민통치를 방조하는 주체로 따가운 시선을 받고, 이로 인해 미묘한 입장에 처할 수밖에 없던 모순적 상황으로 인해 중국인과의 갈등이 불가피했을 당시 조선 이주민의 현실을 떠올려 볼 수 있다. 계급주의적 관점이 뚜렷하게 드러나는 <만보산>에 비해 상대적으로 <농군>에서 민족주의적 시각이 강하게 드러나는 원인은, 바로 이러한 지점과 연계해서 이해할 수 있는 것이다.

5. 나오며: 디지털 환경에서의 문학작품 연구에 대한 제언

본 연구는 ‘데이터를 활용한’ 문학작품 연구를 표방하였지만 소위 ‘빅데이터’라 불리는 큰 규모의 데이터를 구축하지 않았고, 그러한 큰 규모의 데이터를 활용해 소위 ‘멀리서 읽기(distant-reading)’라 불리는 인간 연구자의 인지 능력을 뛰어넘는 컴퓨팅 기반의 분석을 시도하지도 않았다. ‘문학 연구에서 데이터를 다룬다는 것은 기존의 질적 연구 관점에서 벗어

24) 고명철, 『동아시아 반식민주의 저항으로서 일제말의 ‘만주 서사’ -이태준의 <농군>과 한설야의 <대륙>을 중심으로-』, 『한국문학논총』 제49집, 2008, 한국문화회, 91면.

나 양적 차원에서 어떠한 사실을 발견하고자 하는 노력이 아닌가?’라는 시각에 입각할 경우, 본 연구에서 활용한 방법론과 결과물의 형식은 다소 뜬금없는 것으로 수용될 수도 있다. 그러한 생각은 데이터(data)를 숫자로 바라보는 관점을 전제하는 것이기에, 본 연구에서 강조하고자 하는 데이터의 외연과는 다소 다른 지점에 있다고 할 수 있다.

본 연구에서 강조하고자 한 ‘문학 연구에 있어서 데이터 활용’의 지점은, ‘빅데이터’나 ‘멀리서 읽기’에서 강조되는 인간 연구자의 인지 수준을 벗어나는 계량적 분석으로서 컴퓨팅 방법론의 활용이 아니라, ‘스몰 데이터(small data)’와 ‘꼼꼼히 읽기(micro-reading)’에 기초한 연구자의 문제의식을 반영한 데이터 편찬(compilation)과 공유(sharing)의 실행이라 할 수 있다.

일반적으로 스몰 데이터²⁵⁾는 빅데이터에 비해 그 규모가 작다는 점만 강조되는 경향이 있다. 그러나 비정형 데이터를 보편으로 하는 빅데이터에 비해 상대적으로 규모가 작은 만큼 세밀한 공정을 거침으로써 높은 품질을 갖출 수 있는 것이 스몰 데이터의 특징에 해당한다. 규모는 크지 않지만 품질이 높다는 것은, 곧 상대적으로 고도화된 질적 차원의 과업 수행에서 정교한 접근과 객관성 확보를 위한 최소한의 매개로서 스몰데이터가 기능할 수 있음을 의미한다. 작은 규모를 약점으로 보지 않고 오히려

25) 스몰데이터에 관해서는 다음의 서술을 참고할 수 있다. “데이터는 어떤 주체가 특정한 목표나 의도를 갖고 만들어낸 것으로, 주어진 문제를 해결하기 위해 적절히 활용될 때 그 가치가 발현된다. 물론 주어진 문제가 다차원의 복잡성을 갖고 있다면 충분히 규모 있는 데이터가 효과적이다. 그러나 주어진 문제를 해결하는 데 필요한 최소의 데이터만으로 효과적 결과를 얻는 것도 가능하다. 스몰 데이터(small data)는 규모 중심의 빅데이터 현상과 구별하여 품질이 높은 최소의 데이터를 말한다. 일반적으로 빅데이터는 데이터의 수집과 분석을 위해 시간과 비용이 상당히 소요된다. 스몰 데이터는 특정 문제에 대한 상황을 이해하기 위해 필요한 데이터만을 수집하고, 신속한 분석을 통해 문제를 해결할 수 있는 기회를 제공한다.” -김학래, 『지식그래프』, 커뮤니케이션북스, 2017, 5~6면.

려 높은 품질을 강점으로 볼 경우, 질적 연구가 중요한 문학연구에서 스몰 데이터의 활용 가능성은 무궁무진하다.

‘꼼꼼히 읽기’는 그와 같은 스몰 데이터 프로세싱을 문학 연구에 활용할 경우 시도할 수 있는 연구방법론에 해당한다. 예컨대 품질은 낮지만 큰 규모의 데이터를 활용해 ‘멀리서 읽기’를 시도할 경우, 질적 연구에서는 좀처럼 확인하기 어려운 의미의 지점들이 발견될 가능성이 있고, 이러한 측면이 소위 ‘빅데이터 기반의 읽기’로서 ‘멀리서 읽기’의 유효함을 강조할 수 있는 부분이다. 그에 비해 ‘꼼꼼히 읽기’는 규모는 작지만 높은 품질로 편찬된 데이터를 활용해 말 그대로 꼼꼼하게 텍스트를 독해하는 시도를 가리킨다. 이와 관련해 ‘다루는 대상의 규모에서 변별이 없다면, 아날로그 환경에서 전통적으로 이루어져 온 질적 연구와 비교해 데이터 기반의 꼼꼼히 읽기가 무슨 이점이 있는가’라는 의문을 제기할 수 있다. 여기에 대한 답은 ‘객관적 근거 확보’와 ‘세밀한 정보의 공유’로 압축된다.

질적 연구로서 기존 문학 연구는 소위 ‘글쓰기’를 통해 이루어져 왔다. 글쓰기는 문학 연구자의 논리를 정리하고 드러내는 최적의 미디어로서 여겨져 왔지만, 이른 바 체계적이고 논리적인 과학(science)으로서 문학 연구를 정립하기에는 그 한계가 뚜렷하다. 예컨대 분명한 근거 없이 이루어지는 소위 ‘뭉게기’식의 서술이 문학 글쓰기에서 종종 나타나는 것은 부정하기 어려운 사실이며, 이로 인해 반박과 재반박 그리고 재재반박이 글쓰기를 통해 벌어지는 현상 또한 보편적이다. 문학은 과학이 아니기에, 주정주의적 시각을 드러내는 문학만의 연구 시각과 방법이 있다는 관점에 입각할 경우, 그와 같은 글쓰기의 양상은 문제가 되지 않는다. 그것조차도 문학 연구의 외연으로 포섭할 수 있기 때문이다. 그러나 ‘사실’로서 명확한 근거에 입각한 연구를 수행해야 한다는 학문의 원론적 입장을 고려할 경우, 글쓰기의 한계를 보완할 수 있는 연구 방법론에 대한 고민은 일견 자연스러운 것이다.

본 연구의 세 번째 장 ‘연구방법’에서 제시한 ‘시맨틱 데이터 모델링’ 서술에서 드러나듯이, 연구 데이터를 조직하기 위한 모델링 과정은 작품 연구를 수행하기 위한 최소한의 논리적 체계를 정의하는 것이며, 그렇게 정의된 논리에 입각해 구축된 데이터는 연구를 수행하는 데 객관적 근거로서 활용된다. 네 번째 장 ‘연구결과’에서 제시한 작품 비교의 정확성은 바로 그러한 논리적 체계를 기반으로 구축된 스몰 데이터를 능동적으로 활용한 결과이다. 분석과 해석의 과정에서 연구자의 주관을 최대한 배제하고 객관적 정보를 매개로 작품을 들여다보고자 한 실험적 시도라 해야 할 것이다.

한편으로 작품 분석과 해석의 과정에서 데이터를 활용해 객관성을 확보하고자 하였으나, 애초 데이터를 디자인하는 과정에서 연구자의 주관적 문제의식이 반영된다면 그렇게 해서 구축된 데이터의 활용은 객관성을 잃는 것이라는 주장이 있을 수 있다. 그와 같은 가능성을 부정할 수는 없겠으나, 그 또한 연구 방법에서 제시한 것이기에 만약 데이터 디자인 과정이 잘못되었음을 논증하고자 한다면, 글쓰기가 아니라 본 연구에서 디자인한 데이터 모델링 과정의 잘못된 논리 또는 구축된 데이터의 오류를 지적할 수 있을 것이다. 작품 연구에 있어 데이터의 편찬과 활용은, 그것 자체로 객관성을 확보한다기보다도, 해당 데이터를 매개로 논증의 객관성을 도모할 수 있는 여지가 더욱 분명해지는 효과가 있다고 해야 한다.

편찬된 데이터는 논증의 ‘객관적 근거’로서 다루어질 수 있을 뿐만 아니라 동일한 작품 또는 유사한 범위의 문학 연구를 수행하고자 하는 연구자들에게 ‘세밀한 정보’로서 공유될 수 있다. 본 연구를 통해 편찬된 데이터셋은 본 연구의 온라인 페이지(<http://www.klbksk.com/wiki/index.php/NovelsAboutManbosanIncident>)의 ‘Dataset’ 항목을 통해 공유되고 있으며, 이를 통해 <농군>이나 <만보산>의 서사 구성과 등장인물 및 장소에 관한 세밀한 정보를 얻고자 하는 연구자들은 누구나 해당 데이터

셋을 다운로드받아 자신의 연구에 활용할 수 있다. 연구 결과물의 공공성을 고려할 경우, 논문으로 발표되는 글과는 다른 차원에서 연구 데이터 공유의 효용을 생각해볼 수 있는 것이다.

본 연구에서 시도한 데이터 기반의 작품 해석 방법론이 기존 문학 연구에서는 다루지 않은 낯선 성격의 것이기에, 그것을 설명하기 위한 서술이 결론부에서 이와 같이 길어졌음을 밝히는 바이다. 낯선 한편 나름 엄밀한 연구 형식을 갖추고 있음을 강조하고자 하였으나, 본 연구는 실험적 시도로서 일종의 시론(試論)이라는 것을 분명히 밝히고자 한다. 시론이기에 다듬지 못한 일부 거친 서술은 부득이한 차원에서 수용될 수 있기를 부탁드리며, 추후 시맨틱 데이터 편찬 방법론을 활용한 ‘스몰 데이터’ 기반의 ‘꼼꼼히 읽기 작업’이 한국문학 연구에서 꾸준히 이루어지기를 기대해본다. 본 연구자 또한 그 과정에서 지속적인 고민을 통해 본 연구에 적용한 연구 방법론을 더욱 다듬어 나갈 계획이다.

참고문헌

1. 기본자료

- 李輝英, 『万宝山』, 『東北流亡文學史料研究叢書·万宝山』, 春風文藝出版社, 2019, 236~374면.
- 이태준, 『농군』, 文章社, 1939. (국립중앙도서관 원문 이미지 제공 서비스)

2. 논문 및 단행본

- 고명철, 『동아시아 반식민주의 저항으로서 일제말의 ‘민주 서사’ -이태준의 <농군>과 한설야의 <대륙>을 중심으로-』, 『한국문학논총』 제49집, 한국문학회, 2008, 83~109면.
- 김병준·천정환, 『박사학위 논문(2000~2019) 데이터 분석을 통해 본 한국 현대문학 연구의 변화와 전망』, 『상허학보』 제60권, 상허학회, 2020, 443~517면.
- 김준현, 『한국의 문학지식 장에서 ‘만보산 사건’이 기억되어 온 몇 가지 방식』, 『한국문학연구』 제51호, 동국대학교 한국문학연구소, 2016, 39~73면.
- 김창호, 『동아시아 ‘타자’ 형상 비교 연구- 만보산 사건을 수용한 한중일 소설을 중심으로』, 『중국현대문학』 제31호, 한국중국현대문학학회, 2004, 379~404면.
- 김철, 『몰락하는 신생(新生) -‘만주’의 꿈과 『농군』의 오독(誤讀)』, 『상허학보』 제9권, 상허학회, 2002, 123~159면.
- 김학래, 『지식그래프』, 커뮤니케이션북스, 2017.
- 류인태, 『데이터로 읽는 17세기 재지사족의 일상 -<지암일기(1692-1699)> 데이터베이스 편찬 연구-』, 한국학중앙연구원 한국학대학원 박사학위논문, 2019.
- _____, 『데이터 기반의 고전 읽기 교육 -『논어』를 대상으로 한 디지털 인문학 강의 사례를 중심으로』, 『인문논총』 제78권1호, 서울대학교 인문학연구원, 2021, 43~73면.
- 박영석, 『萬寶山事件研究 : 日帝 大陸侵略政策의 一環으로서의』, 아세아문화사, 1978.
- 서재현·김병준·김민우·박소정, 『멀리서 읽는 “우리” - Word2Vec, N-gram을 이용한 근대 소설 텍스트 분석』, 『대동문화연구』 제115집, 성균관대학교 대동문화연구원, 2020, 349~386면.
- 윤해동, 『“만보산 사건”과 동아시아 ‘기억의 터’ -한국인들의 기억을 중심으로-』, 『사

- 이(SAI)』 제14호, 국제한국문학문화학회, 2013, 479~514면.
- 이재연, 「작가, 매체, 네트워크 -1920년대 소설계의 거시적 조망을 위한 시론-」, 『사이(SAI)』 제17호, 국제한국문학문화학회, 2014, 257~301면.
- _____, 정유경, 「국문학 내 문학사회학과 멀리서 읽기 -새로운 검열연구를 위한 길마중」, 『대동문화연구』 제111집, 성균관대학교 대동문화연구원, 2020, 295~337면.
- _____, 「'문학 동인의 시대' 전후: 한국 작가 네트워크의 통시적 분석 1917-1927」, 『공학으로 인문학 읽기』, 세창출판사, 2021, 75~117면.
- 장문석·류인태, 「디지털 인문학과 한국문학 연구(1) : 작가 연구를 위한 시맨틱 데이터베이스 설계」, 『민족문학사연구』 제75권, 민족문학사연구소, 2021, 347~426면.
- 장영우, 「민족 갈등에서 계급 투쟁으로의 전화(轉化) -이회영, 『만보산(萬寶山)』론-」, 『한국문학연구』 제51호, 동국대학교 한국문학연구소, 2016, 107~140면.
- 전성규·허예슬·최주찬, 「이인직 소설의 감정론 :시각화와 계량화를 기반으로」, 『민족문학사연구』 제76권, 민족문학사연구소, 2021, 135~189면.
- 조은경·드프레메리웨인, 「『별건곤』의 텍스트 근대성 -데이터과학적 언어분석」, 『문화와 융합』 제43권10호, 한국문화융합학회, 2021, 35~54면.

Abstract

A Data-based Comparative Research on Korean and Chinese novels

- Based on Lee Taejun's <The Farmer>
and Li Huiying's <Mt. Manbo> -

Li, Ze · Ryu, In-tae

This research created semantic data for two novels, <The Farmer> by Lee Taejun and <Mt. Manbo> by Li Hueing. We attempted to investigate the possibilities of fresh research using the semantic data compilation methodology by implementing the created data as a knowledge graph. The Manbosan Incident is a Korean, Chinese, and Japanese historical event. Because of their qualities that reflect the perspectives of each nation, Korean, Chinese, and Japanese novels based on the incident have been the subject of continual discussion. Meanwhile, a number of scholars have sought to view the works in a comparative light. However, except from the conclusion based on a qualitative point of view, there has been no study detailing the literary distinctions linked across novels. In order to compare the topography of the entire work, a deductive standard is required, as well as an inductive dimension review based on it. By utilizing the triple(S-V-O) format data compilation mutatis mutandis, the semantic data borrowed from this research is focused on precisely defining the varied range of humanities knowledge. One of the data-based non-linear reading approaches given as the so-called 'digital humanities' is this approach. In the process of reading texts, this allows for the creation of deductive norms and the execution of inductive evaluations. This research will have ramifications as an experimental search for innovative reading using semantic data compilation and knowledge graph implementation in a context where research on works utilizing data analysis methods has recently increased.

【Key words】 modern novel, *Manbosan Incident*, Lee Taejun, Li Huiying, <*The Farmer*>, <*Mt. Manbo*>, digital humanities, semantic data

논문 투고: 2022년 03월 15일, 심사 완료: 04월 10일, 게재 확정: 04월 12일.