

## 문학 작품의 거리 측정을 활용한 야담의 이본 연구

-〈옥소선 이야기〉를 대상으로

권기성\* · 최운호\*\* · 김동건\*\*\*

### 〈차 례〉

1. 서론
2. 전체의 장: 연구대상과 방법론 점검
  - 1) 연구대상과 선행연구
  - 2) 연구방법론
3. 〈옥소선이야기〉의 비교 결과와 그 의미
  - 1) 〈옥소선 이야기〉의 계열 분류의 실제와 점검
  - 2) 기문총화계열 이본 간의 차이와 새로운 문제점
4. 결론

### 〈국문초록〉

본 연구는 디지털 도구를 이용하여 문학 작품의 거리 측정을 계산하고, 이 결과를 통해 〈옥소선 이야기〉 이본의 계열 분류를 시도해 본 것이다. 그간 〈옥소선 이야기〉의 전개과정이나 변이양상에 대한 시도는 수 차례 있었던바, 본 연구는 이를 부정하거나 새로운 측면을 밝히기보다는 야담 이본연구의 새로운 방법론을 도입해 보고자 하였다. 이에 기존 연구를 객관적으로 증명하고, 새로운 가능성을 밝히는 데 초점을 두었다.

우선 문자 단위 서열 비교와 단락별 의미유형의 비교를 통해, 이본 간의 거리를 측정하고 이 결과를 시각화하여 기존의 논의를 점검하였다. 살펴본 바와 같이, 〈옥소선 이야기〉의 계열분류는 선행연구의 지적과 유사했으며, 차이점이 발생하는 원인 역시 야담사의 전개과정에서 해석될 수 있는 부분이었다. 다만 공식적 차원에서 동일

\* 제1저자, 창원대학교 국어국문학과 조교수

\*\* 공동저자, 목포대학교 국어국문학과 부교수

\*\*\* 교신저자, 경희대학교 후마니타스칼리지 교수

계열내의 이본연구는 상대적으로 미흡하다는 것을 알 수 있었다. 이는 서사적으로 유의미한 차이가 발생하지 않기 때문일 텐데, 그럼에도 야담사에서 큰 비중을 차지하고 있는 기문총화계열의 객관적이고 정량적인 연구가 필요함을 알 수 있었다.

단일 이야기가 아니라 다수의 야담집을 연구 대상으로 삼을 경우, 인간의 처리능력으로는 필연적 실수와 비효율적 품을 들이게 된다. 그런데 지금껏 야담의 이본연구는 대부분 전통적인 인문학의 방법으로 진행되어 왔다. 그것은 그것대로 인정하면서, 새로운 가능성을 모색하는 시도도 동시에 이루어져야 할 것이다. 이는 향후 야담 향유양상의 실제적 구획과 관계양상 등을 시각화시키는 작업으로 이어질 수 있으며, 디지털 인문학이라는 새 시대의 연구방법론을 활용하여 야담 이본연구의 지평을 확장할 수 있을 것이다.

옥소선 이야기, 이본 연구, 전승과 변이, 계열과 계통, 디지털 인문학

## 1. 서론

본 논문은 문학 작품의 거리 측정을 활용하여 야담집에 수록된 〈옥소선 이야기〉의 이본 연구를 수행하는 데 그 목적이 있다.

근자에 디지털 인문학이라는 키워드는 문학 연구의 새로운 시도로 이루어지고 있다. 개념 정의에 대한 합의를 잠시 차치하더라도,<sup>1)</sup> 연구 대상의 시대와 장르를 불문하고 다양한 문학 작품의 연구방법론으로 사용되고 있는바, 그 결과물이 공히 집적되었음은 자명한 사실이다.<sup>2)</sup> 새로운 시대는

1) 본고에서 사용하는 디지털 인문학이라는 개념은 다음의 정의와 유사하다. “단순히 인문학 자료를 디지털화 하거나, 연구 결과물을 디지털 형태로 간행하기보다는, 정보 기술의 환경에서 보다 창조적인 인문학 활동을 전개하는 것을 디지털 인문학이라고 한다. 종래 인간의 언어로만 기술되었던 인문 지식의 요소들을 기계가독적 데이터로 전환하여 컴퓨터가 그 정보의 해석을 도울 수 있게 하는 것이라고 할 수 있다.” 김현·안승준·류인태, 「데이터 기반 인문학 연구 방법의 모색 : 문중 고문서 아카이브와 디지털 인문학의 만남」, 『횡단인문학』 창간호, 숙명인문학연구소, 2018, 19~20쪽.

새로운 인간의 삶을 야기하게 마련이고, 그에 따라 문학을 바라보는 인간의 ‘연구방법론’ 역시 나름의 모색을 피하는 것은 자연스러운 일이다. 그런 까닭으로 위와 같은 현상 또한 이해할 수 있겠다.

다만, 어떤 연구방법론이라도 절대적인 해결책이 될 수 없듯, ‘디지털’이라는 접근 방식 또한 문학해석의 만능열쇠는 될 수 없을 것이다. 일각에서는 ‘디지털’이라는 용어가 마치 인간의 삶과 문제를 전부 해결해 줄 것처럼 이해하곤 하는데, 그러한 오해가 문학 연구에도 쉽사리 통용될 것처럼 여기는 것은 무척 곤란한 일이다. 백번 양보하여 대중적이거나 상업적인 목적의 활용은 그러한 가능성에 기댈 수 있겠으나, 학문적 엄밀성과 객관성을 담보해야 하는 문학 연구는 ‘디지털’이라는 도구를 전가의 보도처럼 사용할 수 없음은 물론이다. 이것이 어디까지나 새 시대의 새로운 연구방법론이라면, 현 단계의 ‘디지털 인문학 방법론’이란 개념은 텍스트 분석을 위한 적절한 도구, 혹은 그 결과물이 기존의 연구사에 유의미한 의의를 더할 수 있다는 가능성으로 접근하는 편이 오히려 유용할 것이다. 다시 말해 이는, 개별 장르와 작품의 성격에 따라 디지털 도구를 활용하는 방식을 더욱

- 
- 2) 디지털 인문학이라는 관점에서 산출된 연구결과물은 대단히 많다. 여기에서 그 모두를 제시할 수는 없기에, 고전문학과 관련된 대표적인 연구업적들만 나열하도록 한다. 최운호·김동진, 「컴퓨터를 이용한 고전문학 디지털콘텐츠의 유사도에 따른 이본 계통 분류 연구」, 『한국정보기술학회지』 12(7), 한국정보기술학회, 2014, 210~213쪽; 권혁래, 「디지털 인문학과 고전문학의 만남」, 『열상고전연구』 50, 열상고전연구회, 2016, 9~12쪽; 강우규·김바로, 「소현성록 연작의 문제론적 고찰 -컴퓨터를 활용한 계층분석을 바탕으로」, 『인문과학연구』 59, 강원대학교 인문과학연구소, 2018, 29~46쪽; 김현·안승준·류인태(2018), 위의 논문, 29~46쪽; 최운호·김동진, 「컴퓨터 문헌 분석 기법을 활용한 〈토끼전〉 이본 연구」, 『우리문학연구』 58, 우리문학회, 2018, 123~154쪽; 김성문·김바로, 「디지털 분석 기법을 활용한 시조문학 연구」, 『문화와 융합』 41(5), 한국문화융합학회, 2019, 747~772쪽; 류인태, 「데이터로 읽는 17세기 재지사족의 일상 : 『지암일기(1692~1699) 데이터베이스 편찬연구』, 한국학중앙연구원 박사학위논문, 2019, 1~275쪽; 박경우, 「머신러닝을 활용한 고시조 연구 방법론 탐색 -운을 패턴 추출 방안을 중심으로-」, 『열상고전연구』 74, 열상고전연구회, 2021, 37~68쪽.

적절하고 치밀하게 설계해야 함을 의미한다.

이러한 문제의식 하에 본고는 고전문학의 서사장르인 야담집 소재 작품을 대상으로, 문학 작품의 거리 측정을 활용한 이본 연구를 수행하는 데 초점을 맞춘다. 그 이유는 다음과 같다. 야담은 조선후기 사회현실의 변화를 생생하게 담아내며 성장한 한문서사의 일종이다. 이러한 야담은 당대의 문인들에 의해 채록·편찬되었고, 구전전승 되거나 ‘야담집’이라는 책의 형식을 빌어 전해지곤 했다. 그 외중에 어떤 이야기들은 서사적 필치를 더해 ‘한문단편소설’의 경지로 나아가기도 했다. 야담을 이해하는 데 있어 위의 ‘전승’ 개념은 매우 요긴하다. 실제 다수의 이야기는 시대를 넘어 여러 곳의 야담집에 유사한 형태로 반복 수록되기 때문이다. 편찬자에 따라 ‘전재(轉載)’라 할 만큼 원형태를 유지하기도 했고, 일정한 변이를 거치며 이야기의 변화 양상을 적극 표방한 경우도 있었다. 때문에 야담 연구의 한 축은 여러 야담집에 산견 되는 이야기의 동이(同異)를 살피는 쪽으로 집중되었고,<sup>3)</sup> 그 성과가 상당수 집적되어 많은 이야기의 전승 과정이 소상히 밝혀진 바 있다.

그런데 위와 같은 원본비평 연구를 인간이 수행하기에는 적지 않은 노고와 어려움이 따른다. 예컨대 A본과 B본의 일대일 대응이라면 명쾌한 차이가 드러나겠지만, 작품의 수가 많아질수록 모든 작품의 상대적 차이를 비교하는 일은 꽤 지난한 작업이 된다. 대략적인 구분 내지는 큰 특성에 따른 유형 분류는 가능하겠지만, 이 또한 연구자의 직관에 의지하게 되거나 글자 단위의 세세한 차원에서 오류를 내재할 위험을 상시적으로 안고 있다. 이른바 전통적 연구방법론으로는 많은 시간과 공력이 소요될 뿐 아니라

3) 대표적인 논문을 제시하면 다음과 같다. 정명기, 「홍순언 이야기의 갈래와 그 의미」, 『동방학지』 45, 연세대학교 국학연구원, 1984, 301~341쪽; 이강옥, 「야담집에서의 이경류 이야기의 전개와 그 의미」, 『한국문학논총』 61, 한국문학회, 2012, 61~93쪽; 이채경, 「18세기 야담의 유사 화소 교착 및 변개 양상」, 『대동한문학』 40, 대동한문학회, 2014, 399~441쪽; 강혜규, 「문유채 서사의 변이 양상과 의미」, 『Journal of Korean Culture』 52, 한국어문학국제학술포럼, 2021, 439~465쪽.

다수의 자료를 객관적으로 처리하기가 어렵다는 것이다. 이에 본 연구는 야담의 이본에 따른 전승과정을 밝히는데 있어 좀 더 효율적이고 객관적인 방법론을 모색해보고자 한다. 이를 위해 문학 작품의 거리 측정을 활용, 이본 간 유사도를 직접 비교하고 그 차이를 바탕으로 이본의 계통과 계열을 살필 것이다.

다만 본 방법론은 야담의 이본 연구에 처음 접목하는 것인 만큼 다소 시론적 성격을 띠고 있다. 이에 완전히 새로운 자료를 대상으로 하기보다는 선행연구가 있는 기왕의 작품을 중심으로 하는 편이 보다 효율적이라 판단하였다. 요컨대 본 논문에서는 기존 연구의 정확성을 객관적으로 증명·점검하고 새롭게 밝힌 점을 제시하는 한편, 야담 이본 연구의 새로운 가능성을 제시할 것이다.

## 2. 전제의 장: 연구대상과 방법론 점검

### 1) 연구대상과 선행연구

주지하였듯 본 연구는 기왕의 방법론을 점검하고 객관적으로 증명하는데 1차 목표가 있다. 이에 야담의 대표적 사례라 할 수 있는 <옥소선 이야기>를 대상으로 연구를 진행하고자 한다. <옥소선 이야기>에 대한 연구는 다양하게 전개되어 온 바 있다. 다만 해당 작품의 이본 연구를 수행하고 이를 바탕으로 계통을 분석한 연구는 이신성에 의해 최초로 시도되었다.<sup>4)</sup> 이신성은 12종의 <옥소선 이야기>를 ① 『동패략송』·『계서야담』 계열, ② 『청구야담』 계열, ③ 『천예록』·『동야회집』 계열로 나누었다. 장진숙<sup>5)</sup>은

4) 이신성, 「〈옥소선이야기〉의 전개양상과 그 의미」, 『파전김무조박사화갑기념논총』, 동간행위원회, 1988, 1~17쪽.

5) 장진숙, 「〈옥소선 이야기〉의 전변양상과 그 서사적 의미」, 『고전문학과 교육』 16, 한국고전문학교육학회, 2008, 297~326쪽.

이러한 분류에 통시적 의미를 부여하였다. 곧 <옥소선 이야기>는 시대의 흐름에 따라 전기적 성격을 띤 야담계 소설 <소설인규옥소선(掃雪因窺玉簫仙): 『천예록』>에서 야담의 전형성을 확보한 야담계 소설 <고유일재(古有一宰): 『계서야담』>, <청기어패자등과(聽妓語悖子登科): 『청구야담』>으로 이어졌으며, 소설적 문체를 지닌 야담계 소설 <소설정획규고정(掃雪庭獲窺故情): 『동야휘집』>의 유전을 보여주고 있다는 것이다. 동일한 유회의 이야기이지만, 그 이야기가 놓인 시대에 따라 다른 면모들을 드러내며 유전되고 기록된다고 하였는데, 자료의 분류 자체는 이신성의 기준과 크게 다르지 않았다.

김준형은 기존의 연구성과를 바탕으로 자료를 추가하여 40개의 자료를 8개의 계통으로 분류하였다.<sup>6)</sup> 그에 따르면 <옥소선 이야기>는 ① 『천예록』 계열, ② 『동패락송』 계열, ③ 『기문총화』 계열, ④ 『동야휘집』 계열, ⑤ 『청구야담』 계열, ⑥ 동양문고본 『청구야담』, ⑦ 고려대본 『월하선전(月下僊傳)』, ⑧ 『월간야담』 소재 「소설정획침고인(掃雪庭獲窺故人)」으로 나뉜다. ①~⑤는 동일 계열 내에서 차이가 대동소이하며, ⑥~⑧까지는 작가에 의한 적극적인 개작이 두드러진다고 하였다. 이 연구는 <옥소선 이야기>에 대한 가장 자세한 분류일 뿐 아니라, 이신성, 장진숙 등의 작업에 비해 다양한 자료들을 검토하여, 해당 이야기의 전승과정을 소상히 밝혔다는 점에서 연구의 의의가 분명하다.

이상의 성과는 주로 자료의 통시적 전개를 통한 ‘서사적 발전과정’을 밝히는 공통점을 보인다. 하지만 선행연구에서도 밝히고 있듯, ‘통시적 측면으로 접근하면서 간과된 공시적인 공존양상’에 대한 분석<sup>7)</sup>은 일괄적으로 거의 진행되지 못했다. 특히 『기문총화』 계열로 분류되고 있는 『계서잡

6) 김준형, 「<옥소선 이야기>의 변이양상과 의미」, 『한국민속학』 30, 민속학회, 1998, 265~288쪽.

7) 장진숙(2008), 위의 논문, 322쪽.

록』, 『계서야담』 등의 상세한 비교는 언급되지 못했는데, 이는 서사적으로 유의미한 차이가 보이지 않았기 때문으로 이해된다. 이에 본고는 기존의 계열 분류의 적합성을 점검하는 한편, 공시적 공존양상을 밝히는 데도 주력하고자 한다. 이러한 목적 하에 본고에서 다루는 <옥소선 이야기>의 전체 목록을 제시하면 아래와 같다.

<표1> <옥소선 이야기> 이본 목록

NO	이 본	야담집명	계열	ID
1	천리대본	천예록	천예록 계열 (CHUN)	CCHN
2	고려대본 백두산기	천예록		GCHN
3	김영복본	천예록		KCHN
4	연세대본	동패락송	동패락송 계열 (DONG)	YDON
5	동양문고본	동패락송		DDON
6	일사문고본	계서잡록속	계서잡록 계열 (KIMN)	IKES
7	정명기본	계서잡록		JKES
8	하버드대본	계서잡록		HKES
9	국립중앙도서관본	계서잡록		NKES
10	가람문고본	계서잡록		GKES
11	장서각본	계서잡록		ZKES
12	연세대4책본	기문총화		YKIM
13	동양문고갑본	기문총화		DKIM
14	동양문고을본	기문총화		EKIM
15	가람문고본	기문총화		GKIM
16	천리대본	기문총화		CKIM
17	국도본	기문총화		KKIM
18	정명기본	기문총화		JKIM
19	천리대본	계서야담		CKEY
20	규장각본	계서야담		KKEY
21	연세대본	계서야담		YKEY
22	연세대본	쇄어	YSEO	
23	장서각본	선언편	ZSUN	
24	연민본	성수총화	YSNG	
25	천리대본	총화	CCNG	
26	서울대본	청야담수	SCYD	
27	한남서림본	동상기찬	HDSK	

28	버클리대본	청구야담	청구야담 계열 (CHNG)	BCGY
29	동경대본	청구야담		DCGY
30	교토대본	청구야담		GCGY
31	대관부립도서관본	동야휘집	동야휘집 계열 (DOYA)	DDNG
32	천리대본	동야휘집		CDNG
33	가람문고본	동야휘집		RDNG
34	장지영본	동야휘집		JDNG
35	서울대본	동야휘집		SDNG
36	동양문고본	청구야담	동양문고본 청구야담 계열(DOCH)	YCGY

기왕에 밝혀진 50여 개의 이본 중 36개의 자료를 목록화하고, 해당 자료는 영문 약자로 표기하였다. 이 중 아단문고본 『동괘락송』과 유재영본·고려대본<sup>8)</sup> 『계서잡록』, 규장각본 『선언편』과 정명기본 『청구기화』, 서울대본 『동국고사』, 경북대 프린트본 『동야휘집』, 고려대본 『동야휘집』은 직접 확인하지 못했고, 규장각본 『청구야담』과 국립도서관본 『동야휘집』은 관련기관 소장 자료를 검토하여도 발견할 수 없었다. 또한 고려대본 『월하선전』과 『월간야담』 소재 「소설정획첩고인(掃雪庭獲覘故人)」, 동아일보 연작과 『병자임진록』 소재 등은 근대에 개작된 한글본인 까닭에 연구의 대상에서 제외하였다. 이는 본고가 문장 단위에서 사용된 글자의 스트링 매칭 작업을 기반으로 하였기에, 한자와 표기 방법이 다른 한글에 대해서는 동시 작업이 어려웠기 때문이다. 또한 상대적으로 근대에 산출된 작품인 만큼 개작의 양상이 전근대 자료의 그것과는 응당 다르게 나타나기 때문이기도 하다. 위의 36개의 자료만으로도 본고가 목적으로 하는 유의미한 작업은 충분히 이루어질 수 있기에, 위의 목록에 따라 논의를 진행하였음을 밝혀둔다.

8) 고려대본 『계서잡록』은 한국고전연구학회 117차 동계학술대회 발표 이후, 토론자인 이승은 선생님께 자료를 제공받았으나, 금번 작업에는 추가하지 못했다. 지면을 빌어 감사드리고, 추후의 작업에 보강할 계획임을 밝힌다.

## 2) 연구방법론

본 연구의 방법론은 크게 두 가지로 대별된다. 하나는 레벤시타인 거리 측정법을 적용하여 대응 단락의 문자 단위의 비교와 정렬(alignment)을 수행함으로써 이본 간의 차이를 거리로 환산하여 계산하는 방법이고, 또 하나는 서사단락에 의미 유형을 인코딩하여 그 값에 따라 이본 간의 거리를 계산하는 방법이다. 두 번에 걸친 비교방법을 사용한 이유는, 야담의 전승과정에서 발생하는 독특한 현상에 기인한다. 야담은 문헌으로 전승되기도 하지만, 구연과정을 통해 전승되기도 한다. 그런데 구연과정의 전승을 통해 정착된 것은 같은 이야기라 하더라도, 기억에 의거하여 작성되기 때문에 표현에는 차이가 있을 수 있다.<sup>9)</sup> 예를 들어 『계서잡록』을 비롯한 『계서야담』, 『기문총화』 등 일련의 작품들은 『동패락송』과 서사구조는 유사하지만, 표현은 전혀 다르게 나타난다.<sup>10)</sup> 따라서 이런 유형의 작품은 단순히 글자 비교를 통해서 살펴볼 경우, 그 차이의 정도가 왜곡되어 해석될 수 있으므로, 단락별 의미유형의 방법을 함께 제시할 필요가 있는 것이다. 두 가지 측정 결과를 결합하여 제시함으로써 이본 간의 차이를 보다 정확하게 파악할 수 있겠다.

우선 각 이본의 문자 단위를 비교한 내용을 살펴보도록 하자. 이본 간의 차이는 다음과 같이 레벤시타인 거리(Levenshtein distance, LD) 측정

9) 기문총화계는 동패락송계와 거의 유사하다. 그러나 분량면에서 기문총화계는 2100餘字로 동패락송계의 1600餘字보다 500餘字가 많으며, 문체면에서도 동패락송계보다 상당히 소설적이다. 그것은 이희평이 『계서잡록』을 편찬할 때 노명화의 『동패락송』을 본 후, 그 서사구조를 '기억'하여 저술한 것이기 때문이다. 김준형(1999), 앞의 논문, 270쪽.

10) 『계서잡록』은 『동패락송』과 서사의 전개 방식은 동일하나 현실적 문맥은 서사적 도구로 간단히 처리하고 있으며, 상황 설정이나 묘사 및 대사의 확장 등이 활발히 일어난다. 이는 당대 이야기판에서 벌어지던 야담의 구연화를 반영한 결과라 할 수 있다. 권기성, 「『계서잡록』 4권 소재 야담의 서사방식 - 『동패락송』과의 비교를 중심으로」, 『한국한문학회연구』 80, 한국한문학회, 2020, 71~105쪽.

법11)을 통해 글자 단위를 대상으로 진행된다.

	1	2	3	4	5	6	7	8	9	10	11	12	
CKIM	與	之	相	狎	恩	情	之	篤	如	山	如	海	12
JKES	與	之	相	狎	恩	情	如	山	女	如	海		11

	1	2	3	4	5	6	7	8	9	10	11	12	13
CKIM	與	之	相	狎	恩	情	之	篤	如	山	-	如	海
JKES	與	之	相	狎	恩	情	-	-	如	山	女	如	海

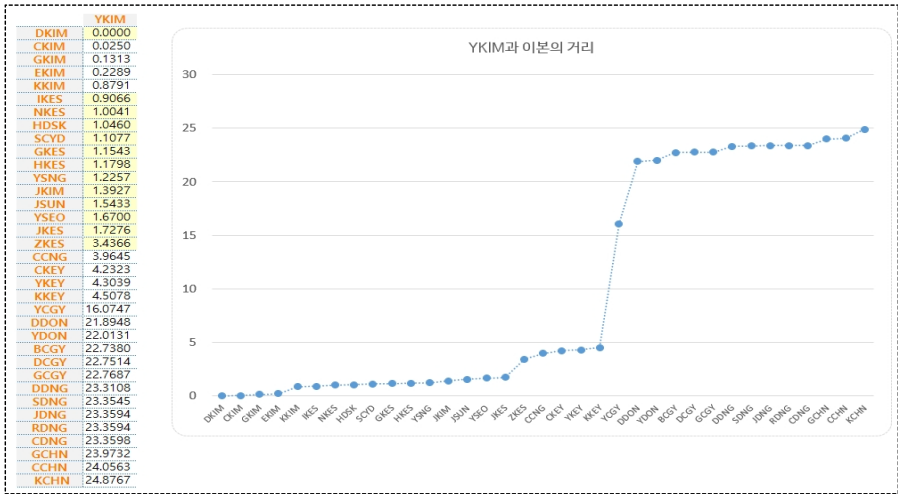
거리(distance)  
 거리(distance) = 3  
 정규화 거리: 3/13 = 0.231

[그림1] 두 이본의 글자단위 정렬과 비교

이들테면 위와 같은 그림을 사례로 살펴보자. CKIM(천리대본 기문총화)와 JKES(정명기본 계서잡록)의 각 구절을 대상으로, 길이가 다른 문장을 서로 같은 라인으로 정렬을 하였다. 요컨대, 첫 번째 문장의 경우, CKIM의 ‘與之相狎恩精之篤如山如海’(12)과 JKES의 ‘與之相狎恩精如山女如海’(11)을 실제 스트링 길이가 아니라 짝을 맞춘 정렬 길이로 정렬한 뒤, 전체 길이를 13으로 설정, 그 안에서 서로 빠지고 바뀐 글자를 비교하는 것이다. 이렇게 보면 두 이본 간에 문장을 이루는 스트링 길이의 차이는 3만큼 있다고 할 수 있고, 정확히는 두 문장의 차이가 3/13(0.231)이 되는 것이다. 두 이본에 해당하는 모든 문장에서 삽입(i : insertion)이든

11) 레벤스타인 거리(Levenshtein distance, LD) 측정법은 두 개의 스트링이 있을 때 스트링 s를 스트링 t로 바꾸는 데 사용되는 비용을 측정하는 방법이다. 비용의 측정에는 삽입(insertion), 삭제(deletion), 대치(substitution) 세 가지 비용을 계산한다. 이 방법은 동적 프로그래밍(dynamic programming) 기법을 사용하여서 원천 스트링(source string)을 목표 스트링(target string)으로 변환하기 위해 소요 되는 비용을 계산함으로써 두 스트링의 유사도를 계산하는 알고리즘이다. 이 방법에 의하면, 어떤 스트링 S와 T의 거리는, S를 스트링 T로 변환할 때 필요한 삽입, 삭제, 대치 연산의 합으로 정의된다. 과정을 동적 프로그래밍 기법으로 구현한 것이 LD 알고리즘이다. 최운호·김동진, 「군집분석 기법을 이용한 텍스트의 계통분석」, 『인문논총』 62, 서울대학교인문학연구원, 2009, 210~213쪽.

삭제(d : deletion)든, 치환(s: substitution)이든 차이가 나는 부분을 1씩 거리(distance)를 늘린다. 이와 같은 방식으로 작품 내 모든 문장을 전체 이본을 대상으로 서열 정렬(sequence alignment)을 진행하고 스트링(시퀀스)의 차이를 계산하면, 필사본 전통에 따른 한문 스트링 비교가 가능해진다.

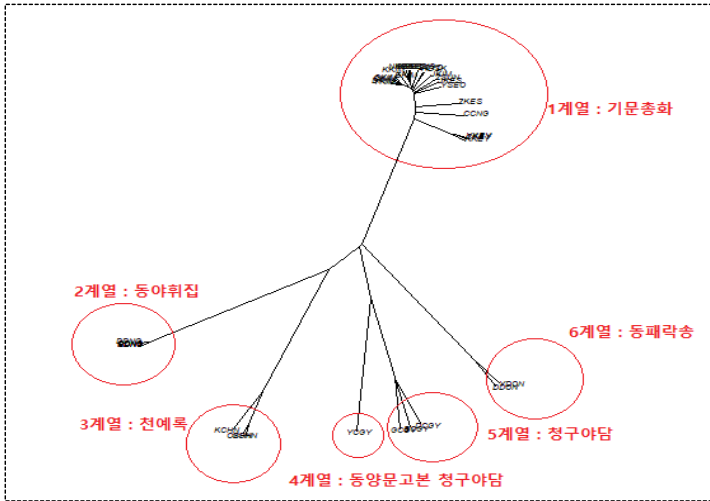


[그림2] YKIM 대비 여타 이본과의 유사도 그래프

사례로 위 그림을 살펴보자. 이는 위의 방법에 따라 YKIM(연세대본 기문총화)의 여타 이본들에 대한 차이를 계산한 것이다. 좌측은 계산 값으로 숫자의 값이 적을수록 유사한 것이며, 우측은 유사도 그래프로, 좌측에서 우측으로 갈수록 차이를 보이는 것이다. 그림을 통해 YKIM(연세대본 기문총화)는 기문총화 계열 혹은 계서잡록 계열 이본들과 가깝고, 청구야담 계열이나 동아회집 계열과는 상대적으로 먼 거리를 보이는 만큼, 더 많은 차이가 발생함을 알 수 있다. 이렇듯 전체 단락의 문장을 1:1로 비교하고 그 결과를 길이로 나눈 뒤 값을 모두 합하면, 전체 이본의 판본 대 판본의 거리가 나오게 되며, 이를 통해 각 문장의 정규화된 거리 차이를 계산한

결과를 파악할 수 있게 된다.

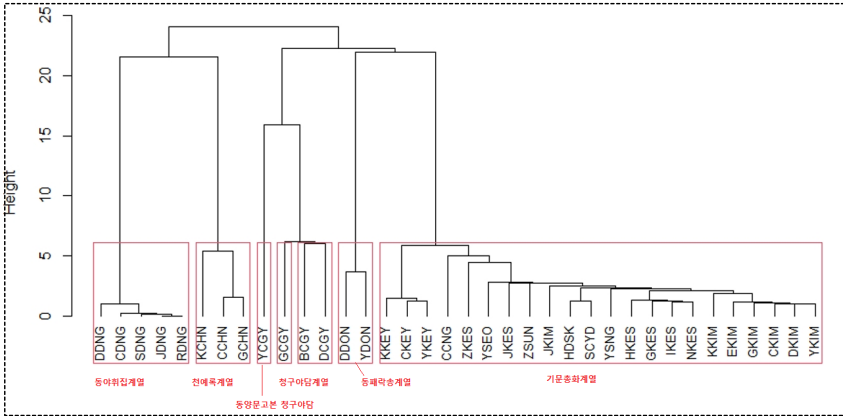
아래의 [그림3]은 이러한 결과값을 계통분기 분석을 통해 시각화 한 것이다. 계통분기란 서로 연관관계를 가진 그룹들을 찾아내고 이들 간의 상호 연관관계를 밝혀내는 작업을 하면서 이들 간의 발생적인 연관관계를 전제하고 그 관계망을 그려내는 작업<sup>12)</sup>을 의미한다. 위 그림은 개별 이본들의 상대적 거리를 보여주고 있기 때문에, 이본 간의 거리가 길면 길수록 연관관계가 약하다는 것을 의미하며, 일정 지점에서 분기하여 군집되어 있는 경우 유사한 계열임을 의미한다. 이를 통해 보면 전체 6개의 그룹으로 묶여 있는 것을 볼 수 있는데, 1) 기문총화 계열, 2) 동아휘집 계열, 3) 천예록 계열, 4) 동양문고본 청구야담, 5) 청구야담 계열, 6) 동패락송 계열로 나뉘어 살펴볼 수 있다. 이러한 결과는 아래 계층적 군집 분석 방식을 통해



[그림3] 문자 단위 비교 결과의 계통 분기 시각화

12) 강성용, 「고전텍스트의 DNA를 묻는다 : 고전텍스트의 문헌전승 계통도 분석에서 계통발생 분석방법(phylogenetical analysis)의 적용과 모색」, 『인문언어』 15(3), 국제언어인문학회, 2013, 100쪽.

서도 동일하게 나타났다.



[그림4] 글자단위 비교의 계층적 군집분석

계층적 군집분석(hierarchical cluster analysis)은 군집 분석의 일종으로, 가까운 거리를 차례로 묶는 군집방법이다. 이 방법은 군집의 형성에 위계를 주어 일단 한 군집에 속하게 된 각각의 개체는 동시에 속할 가능성을 잃어버리게 되는 단점이 있지만, 자료 간 상호관계를 덴드로그램(dendrogram)으로 표현하기 때문에 자료의 성격을 직관적으로 이해할 수 있게 한다.<sup>13)</sup> 이를 통해 유사한 속성의 이본들을 한 곳에 군집하는 양상을 정확히 파악할 수 있다. 계층적 군집분석을 통해 시각화 된 그림을 보면, 그룹핑 된 계열의 가지가 6개로 위의 계통분기분석과 동일하다. 다만 이 그림을 통해서 각 이본명을 정확하게 파악할 수 있는 바, 개별 작품은 유사한 이본별로 묶여 있음을 알 수 있고, YSEO(연대본 쇄어), ZSUN(장서각본 선언편), YSNG(연민본 성수총화), CCNG(천리대본 총화), SCYD(서

13) 우숙영 외, 「상대적 계층적 군집 방법을 이용한 마이크로어레이 자료의 군집분석」, 『한국데이터정보과학회지』 25(5), 한국데이터정보과학회, 2014, 1000쪽.

올대본 청야담수), HDSK(한남서림본 동상기찬) 등 단일제명의 작품은 모두 기문총화계열로 구분되고 있음을 살펴볼 수 있다.

이번에는 단락별 의미유형에 따라 이본간의 차이를 계산한 결과를 살펴 보도록 한다. 의미유형 비교의 경우, 글자 간 서열 비교가 아니라 단락의 유형에 따라 표지를 인코딩하고, 전체 단락의 표지를 이본별로 계산하여, 그 차이를 비교하는 방법을 사용하였다. 이러한 과정을 위해서는 우선 원시 코퍼스를 서사 단락 단위로 분절하고, 각 서사단락에 계층구조와 내용 유형을 인코딩하여 서사단락유형 주석 코퍼스로 구축할 필요가 있다. 아래 그림은 이런 방식에 해당하는 사례를 제시한 것이다.

ID	CCHN	TITLE	천리대본 천예록
P#1/010/010/000#a#	成廟朝	有一名宰 按節關西 關西自古 以佳麗地擅名 江山樓觀之勝 綺羅管絃之盛 甲于八方 風流豪士 宦遊才子 往往有爲一笑而留三年者	
P#1/010/020/000#a#	妓籍中有小娥一人 名紫鸞 號玉簫仙 年纔巫峽 天賦艷質 絕世無雙 歌舞吹彈 無不精妙 加以才識穎悟 能解詩詞 第一香名 已振關西矣		
P#1/010/030/000#a#	時按使有兒郎 年亦十五 眉眼如畫 幼通經史 藻思敏捷 操筆成章 世以奇童許之 按使無他子女 只有一兒 而才又拔萃 鍾愛特立		
P#1/020/010/000#a#	按使適值懸孤之日 與賓僚置酒於秋香堂 大張妓樂 酒酣歡甚 乃命兒郎起舞 呼首妓 擇於童妓中一人 使之對舞 以爲戲笑之資 衆妓及營中上下 以紫鸞芳姿 妙藝 可敵兒郎 又其年齒 適與同庚 遂使應命 一雙妙舞 嫋嫋如弱柳 翩翩若輕燕 坐上見者 莫不贊嘆 稱其奇絕 按使大悅 招紫鸞 命坐於床頭 饋以肴饌 復以錦綺 厚加賞賚 仍命以紫鸞 永定兒郎陪妓 以供進茶 磨墨之役 自是 恒不移左右 與之同 及數年之後 男女年長 遂相親昵 兩情俱惑 綢繆纏綿 不翅若鄭生之於李娃 張郎之於鶯鶯也		

[그림5] 『천예록』 소재 <옥소선 이야기>의 서사단락유 형 주석 코퍼스

위는 CCHN(천리대본 천예록)의 서사단락 유형 주석코퍼스에 해당한다. <옥소선 이야기>의 첫 부분에 해당하는 위의 단락은 개별 이본마다 각기 다르게 시작되고 있으므로, 동일한 내용 단락을 구분하여 이를 유형화 시켜 표기할 필요가 있다. 이를테면 'P#1/010/010/000#a#'은 해당 단락



만들어진 비교대상 단락 수는 모두 30단락이다. 그러므로 위의 그림은 36종의 이본에 30개의 서사단락마다 인코딩된 유형 표지를 추출하여 유형별로 구성한 데이터를 제시한 것이다.

이본의 유형은 알파벳과 숫자의 조합으로 구성하였다. 작품의 내용 단락 유형에 대한 인코딩 방식은 다음과 같다. 첫째, ‘a’, ‘b’, ‘c’는 전혀 다른 각각의 개별 유형임을 의미한다. 둘째, ‘a1’, ‘a2’, ‘a3’처럼 동일 알파벳에 숫자가 결합되는 경우, 단락 내용의 큰 골격은 동일하지만 변이형으로 존재하고 있음을 의미한다. 셋째, ‘X’는 해당 서사 단락이 탈락하고 없는 경우이며 마지막으로 ‘N’은 해당 서사 단락이 낙장 되고 없는 경우를 의미한다. 이에 따라 36종 이본의 가로축은 30개 서사단락의 세로축에 따라 인코딩 된 유형으로 표지된다.

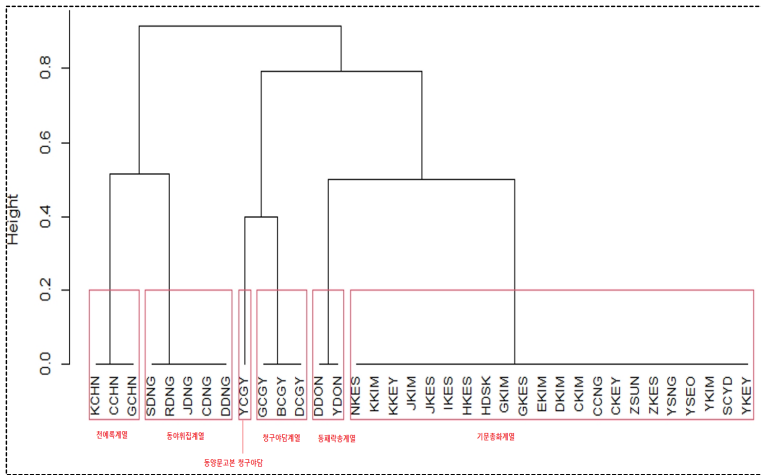
	DKIM	KCHN	BCGY	CCGN	CCNG	CCNG	CKEY	CKIM	CCGY	DDNG	DDON	EKIM	CCGY	GCHN	GKES	GIMM	HDSK	HKES	HES	JONG	JKES	JKIM	
DKIM	0	24.8767	22.738	24.0563	3.96448	23.3598	4.2323	0.025	22.7514	23.3108	21.8948	0.228939	22.7687	23.9732	1.15425	0.131278	1.04598	1.17983	0.90657	23.3594	1.72762	1.39271	—
KCHN	24.8767	0	26.3998	4.3951	24.6253	22.3497	25.0896	24.8767	26.406	22.1579	25.2789	24.9001	26.5214	4.45474	24.7839	24.9001	24.9122	24.8109	24.8292	22.3366	24.8505	25.0455	—
BCGY	22.738	26.3998	0	29.5988	22.749	25.1722	22.8527	22.7411	0.049667	25.1837	21.6431	22.7288	0.185679	25.9458	22.7964	22.7357	22.6556	22.7791	22.7765	25.1688	22.6783	22.8061	—
CCGN	24.0563	4.3951	29.5988	0	23.721	21.2523	24.1511	24.048	25.9546	21.0611	24.7332	24.65	26.0711	0.548611	23.9358	24.0615	24.0307	23.9567	23.9785	21.2444	24.0724	24.312	—
CCNG	3.96448	24.6253	22.743	23.721	0	23.4349	6.63516	3.94038	22.7454	23.389	21.9283	3.94047	22.7665	23.7785	23.5144	3.91254	4.14996	3.7754	3.52561	24.4345	4.28343	4.16215	—
CCNG	23.3598	22.3497	25.1722	21.2523	23.4349	0	23.4979	23.3539	25.1676	10.4467	24.5619	23.3598	25.1369	21.2148	23.3436	23.3562	23.3964	23.3627	23.3696	2.61389	23.4383	23.6469	—
CKEY	4.2323	25.0896	22.8527	24.1511	6.63516	24.9719	0	4.20556	22.8455	23.517	19.9924	4.30648	22.8515	24.1139	4.64038	4.2224	4.80753	4.66073	4.49454	23.4994	5.0554	4.67161	—
CKIM	0.025	24.8767	22.7411	24.048	3.94038	23.3539	4.20556	0	22.7545	23.3049	19.8777	0.203939	22.7717	23.9648	1.12986	0.106278	1.07034	1.15514	0.88178	23.3535	1.70262	1.36771	—
CCGY	22.7514	26.406	0.049667	25.9546	22.7464	25.1676	22.8455	22.7545	0	25.1791	21.5572	22.7402	0.188287	25.952	22.8095	22.7491	22.6691	22.7922	22.7896	25.1642	22.6915	22.8195	—
DDNG	23.3108	22.1579	25.1837	21.0611	23.389	10.4467	23.517	23.3049	25.1791	0	24.347	23.109	25.1484	21.0237	23.3009	23.3135	23.3399	23.306	23.3206	0.98856	23.4062	23.5978	—
DDON	21.8948	25.2789	21.6431	24.7332	21.9283	24.5619	21.9924	21.8977	21.5572	24.347	0	21.8583	21.5463	24.7233	21.9589	21.7997	21.7642	21.9076	21.8678	24.5381	21.9974	21.9881	—
EKIM	0.228939	24.9001	22.738	24.05	3.94047	23.3598	4.30648	0.203939	22.7402	23.109	21.8583	0	22.775	23.9903	1.19435	0.227582	1.24781	1.20508	0.942266	23.3594	1.52715	1.12767	—
CCGY	22.7687	26.5214	0.185679	26.0711	22.7665	25.1369	22.855	22.7717	0.188287	25.1484	21.5463	22.775	0	26.067	22.8486	22.7741	22.6946	22.8074	22.8151	25.1335	22.7066	22.837	—
GCHN	23.9732	4.45474	25.9458	0.548611	23.7785	21.2148	24.1139	23.9648	25.952	21.0237	24.7233	23.9903	26.067	0	23.8851	24.0018	23.9959	23.8993	23.9149	21.2071	24.0086	24.0737	—
GKES	1.15425	24.7839	22.7964	23.9358	23.8436	4.64038	1.12986	0.28095	23.3009	19.9689	1.19435	22.8486	23.8851	0	1.14997	1.59103	0.473693	0.299321	0.29342	1.42261	1.52617	—	
GKIM	0.131278	24.9001	22.7357	24.0615	3.91295	23.3562	4.2224	0.106278	22.7491	23.3135	21.7997	0.227592	22.7741	24.0018	1.14997	0	1.15057	1.1737	0.915828	23.3558	1.69339	1.39127	—
HDSK	1.04598	24.9122	22.6556	24.0367	1.41996	23.3964	4.80753	1.07034	22.6691	23.3399	21.7642	1.24781	22.6946	23.9959	1.59103	1.15057	0	1.40006	1.38192	23.3962	2.18836	2.01625	—
HKES	1.17983	24.8109	22.7791	23.9567	3.7754	23.3827	4.80673	1.15514	22.7922	23.306	21.9076	2.20506	22.8074	23.899	0.473693	1.1737	1.46006	0	0.315128	23.3624	1.51886	1.53242	—
HES	0.90657	24.8292	22.7765	23.9785	3.52561	23.3896	4.8054	0.88178	22.7996	23.3206	21.8678	0.942266	22.8151	23.9149	0.299321	0.915028	1.38192	0.315128	0	23.3692	1.26311	1.26968	—
JONG	23.3594	23.3366	25.1688	21.2444	23.4345	0.261389	3.9594	23.3535	25.1642	0.98856	24.581	23.3594	25.1335	21.2071	23.3432	23.3558	23.3962	23.3624	0	23.438	23.6465	—	
JKES	1.72762	24.6055	22.6783	24.0724	4.28343	3.44883	4.0564	1.70622	22.8915	21.4062	19.9974	1.75275	22.7066	24.0086	1.42611	1.69339	2.18836	1.51886	1.26311	23.438	0	2.0305	—
JKIM	1.39271	25.0455	22.8061	24.1312	4.16215	23.6469	4.87161	1.36771	22.8151	23.5978	19.881	1.17267	22.837	24.0737	1.39127	1.59127	2.01625	1.52422	1.26968	23.6465	2.0305	0	
JUN	1.54329	24.8598	22.7397	24.0161	4.19529	23.5192	4.6048	1.5186	22.7397	21.8225	1.60828	22.7772	23.9674	1.73894	1.51439	1.60828	1.72009	1.54389	23.5189	2.52365	1.82338	—	
KKEY	4.50782	25.113	22.8429	24.1632	6.01472	23.5887	0.397414	4.48108	22.8357	23.6084	22.0107	4.58179	23.8398	24.1242	4.91059	4.49792	5.0897	4.93598	4.70099	5.37164	4.9642	—	
KKIM	0.879146	24.7965	22.6291	23.9401	3.95962	23.4393	4.46833	0.854146	22.8426	23.3951	21.5239	1.01623	22.6675	23.8777	1.53209	0.91714	1.43736	1.53954	1.26605	23.435	2.0172	1.84883	—
NKES	1.00408	24.7899	22.7836	23.9702	3.57356	23.3802	4.53749	0.979388	22.7967	23.3175	21.8701	1.05978	22.8223	23.9114	0.243449	0.999929	1.46766	0.334141	0.147628	23.3599	1.32835	1.41635	—
RDNG	23.3594	23.3366	25.1688	21.2444	23.4345	0.26117	3.8996	23.3535	25.1642	0.95485	24.581	23.3594	25.1335	21.2079	23.3432	23.3558	23.3962	23.3624	0	23.438	23.6465	—	
SDO	1.10765	23.9131	22.6948	24.0345	4.19515	23.4029	4.8032	1.31169	22.7083	23.3465	21.7577	1.28906	22.7388	23.9977	1.63031	1.31989	1.48797	1.42773	23.4028	2.18024	2.07322	—	
SDNG	23.3545	22.237	25.1672	21.1396	23.4297	0.267105	23.4937	23.3486	25.1626	0.990352	24.5573	23.3545	25.1319	21.1903	23.3383	23.3509	23.3913	23.3574	23.3643	0.15853	23.4328	23.6417	—

[그림] <속소선 이야기>의 유사도 일치율

인코딩된 내용 단락의 비교를 유사도 측정에 의해 제시하고 있는 것으로 해밍거리(Hammin distance)<sup>14)</sup>를 이용하였다. 이를 통해 두 이본 간의 차

14) 해밍 거리란 문자열의 전송 도중 몇 개의 글자에서 오류가 났는지를 측정하기 위한

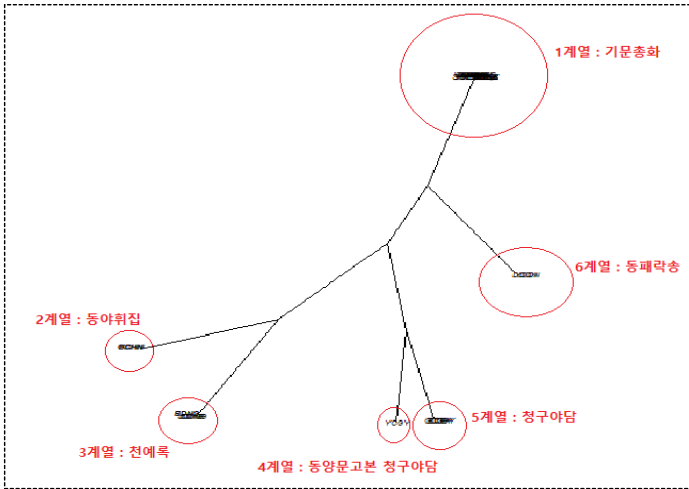
이를 살펴볼 수 있는데, 경우에 따라 가중치를 적용하여 계산 과정을 거침으로써 이본 간 일치율을 수치적으로 나타내어 비교할 수 있도록 하였다. 위의 그림은 그러한 작업의 결과물이다. 단락별 거리는 0~4의 척도로 변환하였다. 동일한 단락은 0, 인코딩된 내용에 유사성이 없는 단락은 2, 동일 내용 단락의 변이는 1의 척도로 거리를 측정하였다. 낙장의 경우에는 계산 대상에서 제외하였다. 이를 앞서 살펴본 계층적 군집분석 및 계통분기 분석으로 제시하면 다음과 같다.



[그림8] 의미유형 결과에 따른 계통분기 분석

의미유형에 따른 분류 결과 계열은 총 6가지로 앞선 문자 단위 비교와 유사하게 나타났다. 다만 비슷한 유형에 대한 동일한 표지가 인코딩 되다

척도로 Hamming(1950)에 의해 소개되었으며, 암호학, 최적화 문제, 정보 이론, 의학 진단 등 다양한 분야에서 많이 활용되고 있다. 안정용, 「의학진단에 이용되는 해밍 거리의 특성 탐색」, 『한국데이터정보과학회지』 23(2), 한국데이터정보과학회, 2012, 231쪽; Hamming, R.W, "Error detecting and error correcting codes", Bell System Technical Journal, 26, 1950, pp.147~160.



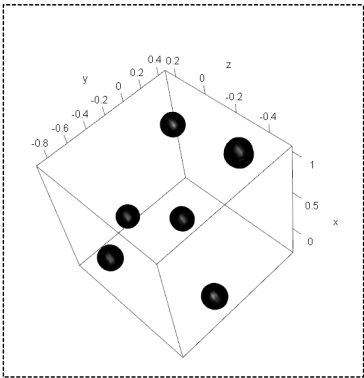
[그림9] 의미유형 결과에 따른 계층적 군집분석

보니, 앞선 문자 단위 서열 비교 정렬 방식에 비해 동일 계열 내에서 이본 군집은 더욱 밀집되고 있음을 살펴볼 수 있었다. 이를테면 기문총화 계열 내의 여러 이본의 밀집도는, 계층적 군집분석의 결과에 의해 나타난 문자 단위 비교에 비해 의미유형 비교가 훨씬 더 집약되어 있는 모습을 볼 수 있다. 이는 기문총화 계열내 여러 이본들이 다소간의 지구차이가 존재하는 반면, 서사 구조나 유형은 거의 동일함을 의미하는 것이다. 한편 위의 두 그림에 따르면 동패락송계열은 기문총화계열의 분기와 가까운 위치에 자리하고 있었는데, 문자 단위 비교에 비해 더욱 근사한 것으로 나타난다. 이는 두 이본계열이 두 작품이 표현의 차이에 비해 비교적 비슷한 의미유형을 공유하고 있음을 의미한다. 다만 두 가지 방법에 따라 계열이 완전히 달라지는 이본은 나타나지 않았다. 이 두 가지 결과를 5:5의 가중치로 보정하여 합한 결과를 3장에 제시하고, 그 결과를 바탕으로 <옥소선 이야기>의 계열분류와 그 의미에 대해 설명하겠다.

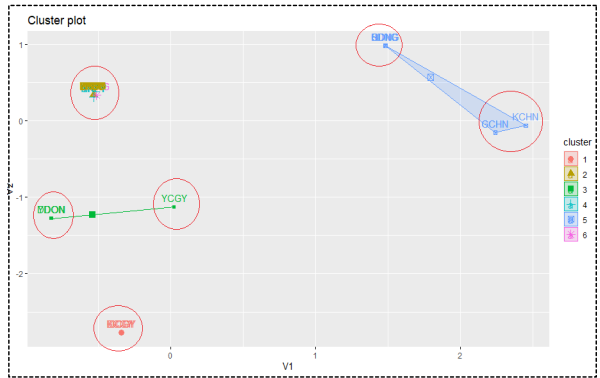
### 3. <옥소선 이야기>의 비교 결과와 그 의미

#### 1) <옥소선 이야기>의 계열 분류의 실제와 점검

본 장에서는 문자 단위 비교와 의미유형 비교에 대한 결과를 바탕으로, <옥소선 이야기>의 계열 분류를 제시하고, 그 의미에 대해 논하도록 한다. 2장에서 살펴본 바, <옥소선 이야기>는 크게 6개의 계열로 분류될 수 있음을 알 수 있었다. 이는 미세한 차이는 있었지만 문자 단위 비교 및 의미유형 비교에서 모두 동일하게 나타났다. 따라서 두 결과를 합친 것 또한 마찬가지로 6가지의 계열로 구분되며, 36개 이본 대상 <옥소선 이야기>의 계열 구분은 6개로 구획된다고 하겠다. 이를 다양한 방법으로 각각 표현해 볼텐데, 이에 따라 판본 유사도와 거리 분석결과를 시각화하여 살펴볼 수 있을 것이다.



[그림10] k-평균법에 따른 이본의 위치

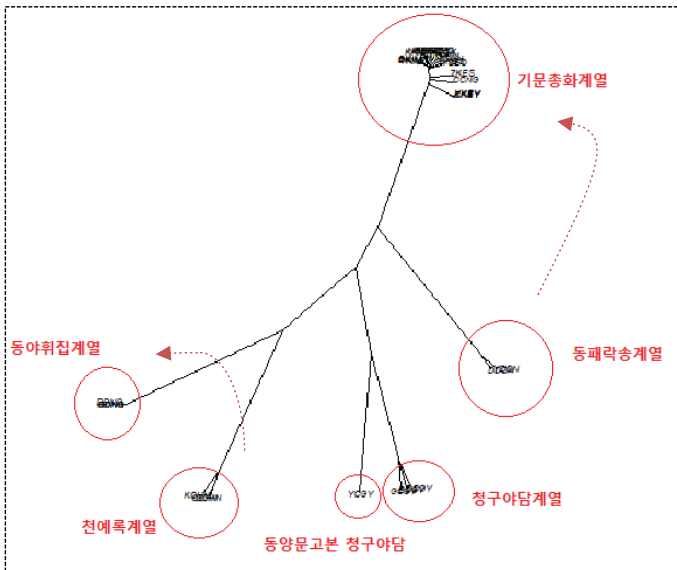


[그림11] 다차원 척도법에 따른 이본의 위치

우선 왼쪽 그림과 오른쪽 그림은 k-평균법과 다차원 척도법<sup>15)</sup>에 따라

15) 다차원척도법(MDS)이란, 다차원 공간에서 정의되는 유사성(similarity) 또는 유사성(dissimilarity)을 저차원의 공간에 기하적으로 나타내어 그들의 관계를 탐색적으로 살펴보는 다변량 그래프적 기법이다. 신상민 외, 「분할법에 의한 가상점을 활용한 다차원척도법」, 『응용통계연구』 28(6), 한국통계학회, 2015, 1171쪽.

각각 <옥소선 이야기>의 이본 군집을 저차원 및 3차원의 공간상에서의 척도에 위치시켜 표현한 것이다. 개별 이본 간의 유사성을 바탕으로 모든 이본 간의 거리를 포지셔닝 맵에 위치시켜 표현한 것으로, 위의 그림을 통해 살펴봐도 <옥소선 이야기>의 계열이 6가지로 분류된다는 사실이 명확하게 드러난다. 이러한 양상은 앞서 살펴보았던 계통분기 분석을 통해서도 알 수 있다. 그러나 이는 어디까지나 작업에 따른 단순 결과물이므로, 해당 계열의 분류가 실상에 맞는지에 대한 점검이 지금부터 필요할 것이다. 계통분기 분석의 그림과 기존 선행연구를 참조해가며, 계열분류의 실상을 살피도록 한다.



[그림12] <옥소선 이야기>의 계통분기 분석을 통한 시각화

그림의 하단 왼쪽 부분부터 살펴보도록 하자. 우선 ①동아휘집 계열과 ②전예록 계열이다. ①과 ②는 동일작품군의 이본을 제외한 모든 판본과

거리가 있다. 특히 기문총화 계열과 가장 먼 거리에 있는데, 그에 비해 두 이본 간의 거리는 상대적으로 가까운 것을 알 수 있다. 『천예록』은 18세기 산출된 이른 야담집이면서도 자저(自著)의 성향이 강하다. 이 내용은 여타의 이본에는 그대로 전제되지 않았고, 19세기 야담집인 『동야회집』과 오히려 가까운 것처럼 나타난다. 그런데 문자 단위 비교를 동야회집(DNG)과 해보면 결핍점이 상당부분 다른 것을 확인할 수 있다. CCHN(천리대본 천예록) 기준으로 동야회집 이본인 CDNG(천리대본 동야회집), DDNG(대판부립도서관본 동야회집), JDNG(장지영본 동야회집), RDNG(가람문고본 동야회집), SDNG(서울대본 동야회집)과는 모두 일정한 수치이상의 차이를 보인다. 그럼에도 다른 야담집에 비해 서로 가까운 것은, 동야회집 계열이 천예록 계열의 서사구조를 참고하는 가운데 만들어진 것임을 알 수 있고, 그 가운데 상당부분이 부연되며 표현의 방식은 다르게 가져갔다고 할 수 있겠다. 즉, 필사전통에서 『동야회집』 편찬자 이원명의 작가적 역량이 많은 영향을 미쳤다고 판단할 수 있을텐데, 『천예록』 원본을 보고 베꼈다고 보다 기억에 의존했거나 새롭게 창작했다고 할 수 있을 만큼 많은 부분을 부연했다고 할 수 있는 것이다.<sup>16)</sup>

서사구조가 동일한데 표현의 방식이 달라지는 관계를 보이는 이본군은, 그림의 상단 중앙과 하단 오른쪽 부분에 있는 ③기문총화계열과 ④동패락송계열을 통해서도 알 수 있다. 『동패락송』은 야담집의 발생순서상 『천예록』의 다음에 나타난 것이다. 『천예록』의 편찬자 임방은 홍봉한의 외조부이며, 『동패락송』의 편찬자 노명흠은 홍봉한 집안에서 반평생 숙사(塾師)로 머물렀기에 이야기를 참고했을 법한데, 두 이본의 거리는 상당히 멀게 나타난다. 반면 『계서잡록』의 편찬자 이희평은 홍봉한 집안의 친인척관계

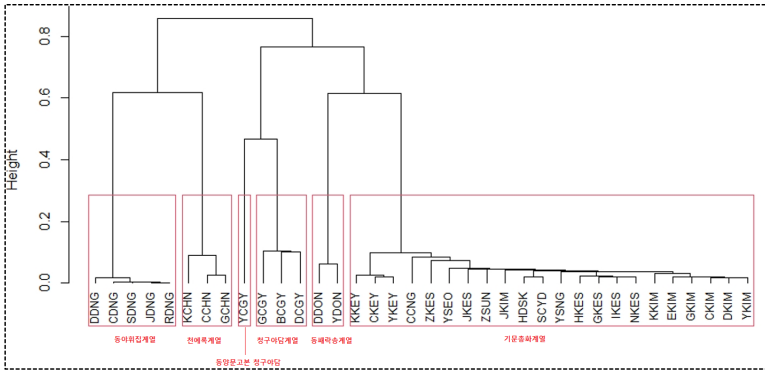
16) 余於長夏調疴，偶閱於野譚紀聞叢話，頗多開眼處。惟是記性衰耗，無以領略萬一，遂就兩書，撮其篇鉅話長，堪證故實者，旁及他書之可資該洽者，並修潤載錄。又采閭巷古談之流傳者，綴文以間之。(『동야회집』序文)

였으며, 『동패락송』을 보았던 정황이 뚜렷하다.<sup>17)</sup> 또한 『계서잡록』으로부터 후대로 이르면서 『기문총화』, 『계서야담』 등이 축출하였을 뿐 아니라, 『선언편』, 『쇄어』 등 유사한 작품들이 이로부터 파생하게 되었다.<sup>18)</sup> 그림을 통해 보면 동패락송계열은 기문총화계열과 분기하여 구분되지만, ‘천예록-동야회집’의 군집보다는 비교적 가깝다는 것을 확인할 수 있다. 그런 점에서 집안 간의 일정한 관련성에도 불구하고, 동패락송계열은 천예록계열보다는 기문총화계열과 가까운 관계에 있다고 하겠다. 그런데 동패락송의 경우에도 문자 단위 서열 정렬을 해보면, 기문총화계열과 상당한 차이가 난다. 대표적인 YDON(연대본 동패락송)을 여타의 야담집과 비교해보면 YKIM(연대본 기문총화)나 IKES(일사문고본 계서잡록) 등 해당 계열의 야담집과 많은 차이가 발생하는데, 이는 동패락송계열-청구야담계열의 차이와 유사한 수준이다. 그럼에도 동패락송이 청구야담계열보다는 기문총화계열과 가까운 것은, 유사한 단락구조를 공유하기 때문인 것으로 여겨진다. 이는 앞서 천예록-동야회집의 경우와 같이 ‘구술-기억-기록’이라는 야담 전승의 영향 때문인 것으로 파악된다.

그림 하단의 중앙에 있는 청구야담계열의 작품들은 비교적 구분이 명확하다. 동일한 분기에서 발생하여 두 개의 기점을 보이는 바, 하나는 YCGY(동양문고본 청구야담)이며, 나머지는 GCGY(교토대본 청구야담), BCGY(버클리대본 청구야담), DCGY(동경대본 청구야담)으로 구획된다. 이는 계층적 군집분석 방식을 통해 보아도 분명하게 드러난다.

17) 이희평의 집안과 홍봉한의 집안은 친인척 관계로 밀접하게 교류하였고, 노명흠의 아들 노궁 역시 홍봉한 집안과 긴밀한 관계에 있었다. 이에 대해서는 다음의 논문들을 참고할 것. 김영진, 「조선 후기 사대부의 야담창작과 향유의 일양상」, 『어문논집』 37, 안암어문학회, 1998, 21~45쪽; 김준형, 「19세기 야담 작가의 존재 양상 - 계서 이희평론」, 『민족문화사연구』 15, 민족문화사학회, 1999, 65~102쪽.

18) 김준형, 「기문총화계 야담집의 문헌학적 연구」, 고려대 석사학위논문, 1997, 1~128쪽.

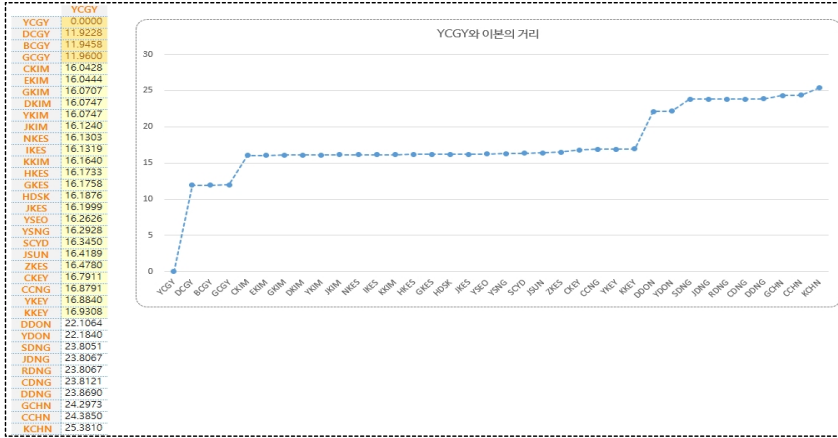


[그림13] <옥소선 이야기>의 계층적 군집분석을 통한 시각화

이 가운데 BCGY(버들리대본 청구야담)과 DCGY(동경대본 청구야담)은 거의 유사한 까닭에 동일한 상관관계에 있음을 확인할 수 있고, GCGY(교토대본 청구야담)은 의미단위로 구분하면 거의 같지만, 문자 단위로 비교하면 미세하게 차이를 보인다. 이러한 현상은 GCGY(교토대본 청구야담) 필사자의 비의도적인 자구 변화에 기인하기 때문으로 생각되며, 서사적으로 유의미한 차이가 있는 것 같지는 않다.

YCGY(동양문고본 청구야담)은 청구야담계열의 군집으로 묶여있지만 나머지 『청구야담』 이본들과 일정한 차이를 갖는다. 아래 문자 단위 서열 비교 결과를 보면 청구야담계열과 가장 유사하면서 기문총화계열과도 유사함을 알 수 있다.

기문총화 계열 중에서도 미세하긴 하지만, 『계서잡록』이나 『계서야담』, 『동패략송』보다는 기문총화류의 작품들과 비교적 더 가까운 것을 살펴볼 수 있다. 이를 통해 보면 YCGY는 청구야담계열의 작품과 『기문총화』 이본을 기반으로 한 합본의 성격을 띠는 것으로 이해할 수 있다.



[그림14] YCGY의 이본 대비 유사도 그래프

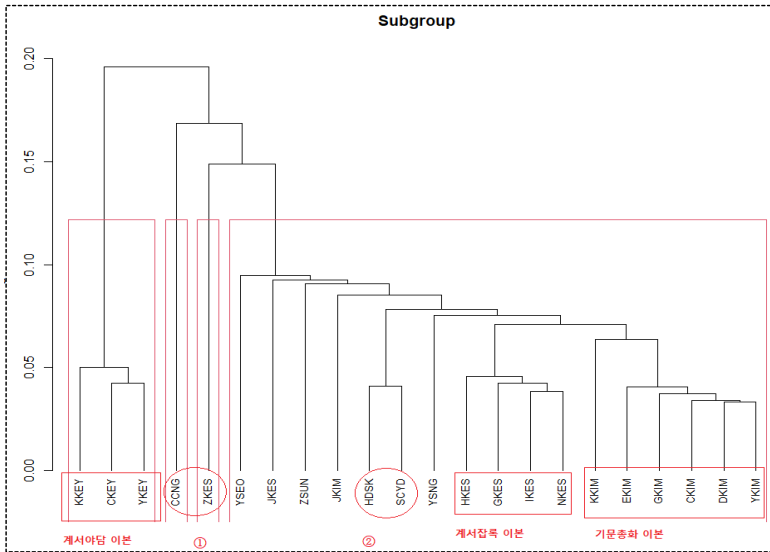
이처럼 <옥소선 이야기>는 크개는 3개, 세부적으로는 6개의 계열로 구분되는 양상을 확인할 수 있었다. 그리고 각 계열별로 나타나는 현상은 기존의 선행연구에서 밝힌 바와 거의 동일함을 확인할 수 있었다. 이러한 결과를 통해 기존 연구가 정확하게 수행되었음을 객관적인 수치로 증명할 수 있게 되었다.

## 2) 기문총화계열 이본 간의 차이와 새로운 문제점

반면 기존의 연구에서는 동일 계열 내의 차이에 대해서는 그리 주목하지 않았다. 물론 이는 동일 제명의 이본군에서 유의미한 지점이 두드러지지 않았기 때문일 터다. 하지만 앞서 살폈던 ‘기문총화계열’ 내의 작품은 다른 자료와는 조금 다른 양상을 띠고 있어 살펴볼 필요가 있다.

우선 이 계열의 작품들은 다른 이본의 군집과 비교하여 꽤 다양한 제명을 갖추고 있는 것이 눈에 띈다. 기존의 연구에 따르면 기문총화계열은 『계서잡록』-『기문총화』-『계서야담』으로 이어지는 야담의 문헌전승을 보여주는 자료로, 동일 제명의 이본도 각기 다양하거니와, 제명을 바꾸었지만

수록 순서나 배치만을 바꾼 야담집이 허다하다.<sup>19)</sup> 이렇게 보면 이 계열의 야담집은 거의 차이가 없는 듯 하지만 밝혀야 하는 점이 없지 않다. 아래의 계층적 군집분석 방식 그림을 보며 따져보도록 하자.



[그림15] 기문총화계열 이본의 계층적 군집분석을 통한 시각화

우선 <옥소선 이야기>의 기문총화계열 내 분류는 크게 두 가지로, 그림의 왼쪽 부분에 『계서야담』 이본이 모여있고, 그림의 오른쪽 부분은 『계서잡록』 이본과 『기문총화』 이본을 위시한 다수의 자료들이 군집되어 있음을 볼 수 있다. 이렇게 볼 때 『계서야담』과 『계서잡록』의 거리는 『기문총화』와 『계서잡록』의 거리보다 상대적으로 멀다는 것을 알 수 있다. 요컨대 『계서잡록』과 『기문총화』의 관련성이 『계서야담』에 비해 더 근사하다는

19) 현재까지 조사된 목록으로 『계서잡록』 35종, 『기문총화』 12종, 『계서야담』 6종 뿐 아니라 제명을 달리하는 야담집까지 포함하면, 그 수는 수십종에 이른다.



한편 ②의 HDSK(한남서림본 동상기찬)와 SCYD(서울대본 청야담수)는 모두 20세기 초 이후 산출된 근대의 작품이다. HDSK(한남서림본 동상기찬)는 1918년 한남서림에서 출간된 한문본이고, SCYD(서울대본 청야담수)는 20세기 초에 산출된 한문현토본이다. 두 야담집의 표기수단에는 차이가 있지만, 실제 양상은 거의 동일한 것으로 나타난다. 두 자료의 경우 문자 단위 서열 비교를 통해 살펴보면, HDSK(한남서림본 동상기찬)과 SCYD(서울대본 청야담수)는 기문총화계열 중에서도 KIM(기문총화)의 이본과 가장 흡사하며, KEY(계서야담)의 이본과 가장 거리가 먼 것으로 파악된다.

	HDSK	SCYD	DKIM	YKIM	CKIM	GKIM	EKIM	IKES	KKIM	HKES	NKES	GKES	YSG	JKIM
HDSK	0	0.232041	1.04598	1.04598	1.07034	1.15057	1.24781	1.38192	1.43736	1.46006	1.46766	1.59103	1.64034	2.01625
	JSUN	YSEO	JKES	ZKES	CCNG	CKEY	YKEY	KKEY	YCGY	DDON	YDON	BCGY	DCGY	CGGY
	2.06058	2.10074	2.18836	3.62363	4.14996	4.80753	4.85948	5.0897	16.1876	21.7642	21.8805	22.6556	22.6691	22.6946
	DDNG	SDNG	JDNG	RDNG	CDNG	GCHN	CCHN	KCHN						
	23.3399	23.3913	23.3962	23.3962	23.3964	23.9959	24.0367	24.9122						

[그림17] HDSK의 이본 대비 유사도 비교 결과

물론 기문총화계열 내의 이러한 차이는 사실 서사적으로 유의미한 변별이라 하기는 어렵고, 또 그 차이도 지극히 미미한 것이 사실이다. 또한 개별 이야기의 유사성을 곧바로 야담집간의 연관성으로 등치시킬 수는 없다. 다만, 그간 야담의 이본 연구가 의미있는 것들의 변화양상에 주목하며 통시적 전개과정을 주로 살폈다면, 공식적인 영향관계에 있는 이본 간의 연구를 통한 동일 계열내의 정밀한 분석이 필요하다는 것을 위의 자료들은 말해주고 있다. 이것이 단순한 글자의 출납 정도의 변화라도, 이러한 작업이 전체 야담으로 수반될 때, 야담집별 차이뿐 아니라, 향유양상 또한 더욱 정확하게 포착될 수 있을 것이다.

때문에 본 작업은 필연적으로 향후 과제를 생각치 않을 수 없게 되었다. 그것은 <옥소선 이야기> 뿐 아니라 기문총화계열이라 불리는 유사 제명의 작품집을 대상으로, 동일 계열 내에서 야담의 지형도를 그리는 작업이 되

겠다. 이러한 작업 또한 기존의 선행논의가 없지 않으나, 본고에서 제시한 야담이본 연구의 새로운 방법론은, 동일 작업을 더욱 객관적이고 효율적으로 분석·처리할 수 있음을 증명하였다.

다만 앞서 언급했듯, 디지털이라는 도구는 만능이 아니라는 점을 다시 한 번 짚어두고자 한다. 그런 점에서 본 연구는 디지털 도구의 ‘신속성’과 ‘정확성’을 활용한 이본의 계통분류 연구에 주목한 것이지, 야담자료의 데이터베이스를 만들거나 모티프의 변이 등을 찾아내고자 한 것이 아님을 밝힌다. 이러한 연구는 또 다른 방법을 이용하여 접근할 필요가 있다.

그간 인간이 수행했던 이본연구야말로, 그 정확성을 점검하기란 꽤 곤란한 일이었다. 본인이 직접 해당 이본의 전체 양상을 살피지 않는 이상 연구 결과를 믿는 수 밖에 없었다. 이런 점에서 본 논문에 제시한 방법론은 차이의 정도와 객관적 거리를 확인할 수 있다는 정확성을 담보하고 있다. 모든 방법론이 그렇듯 일정한 한계를 내포하고 있지만, 디지털 인문학의 시대에 야담연구를 접목할 가능성을 제시했다는 점, 향후 이본 연구에 활용될 수 있는 새로운 가능성을 모색했다는 점에서 나름의 의의를 찾을 수 있지 않을까 한다.

#### 4. 결론

본 연구는 디지털 도구를 이용하여 문학 작품의 거리 측정을 계산하고, 이 결과를 통해 〈옥소선 이야기〉 이본의 계열 분류를 시도해 본 것이다. 그간 〈옥소선 이야기〉의 전개과정이나 변이양상에 대한 시도는 수차례 있었지만, 본 연구는 이를 부정하거나 새로운 측면을 밝히기보다는 야담 이본연구의 새로운 방법론을 도입해 보고자 하였다. 이에 기존 연구를 객관적으로 증명하고, 새로운 가능성을 밝히려는 데 초점을 두었다.

우선 문자 단위 서열 비교와 단락별 의미유형의 비교를 통해, 이본 간의 거리를 측정하고 이 결과를 시각화하여 기존의 논의를 점검하였다. 살펴본 바와 같이, <옥소선 이야기>의 계열분류는 선행연구의 지적과 유사했으며, 차이점이 발생하는 원인 역시 야담사의 전개과정에서 해석될 수 있는 부분이었다. 다만 공시적 차원에서 동일 계열내의 이본연구는 상대적으로 미흡하다는 것을 알 수 있었다. 이는 서사적으로 유의미한 차이가 발생하지 않기 때문일텐데, 그럼에도 야담사에서 큰 비중을 차지하고 있는 기문총화계열의 객관적이고 정량적인 연구가 필요함을 알 수 있었다.

단일 이야기가 아니라 다수의 야담집을 연구 대상으로 삼을 경우, 인간의 처리능력으로는 필연적 실수와 비효율적 품을 들이게 된다. 현대 지금까지 야담의 이본연구는 대부분 전통적인 인문학의 방법으로 진행되어 왔다. 그것은 그것대로 인정하면서, 새로운 가능성을 모색하는 시도도 동시에 이루어져야 할 것이다. 이는 향후 야담 향유양상의 실제적 구획과 관계양상 등을 시각화시키는 작업으로 이어질 수 있으며, 디지털 인문학이라는 새 시대의 연구방법론을 활용하여 야담 이본연구의 지평을 확장할 수 있을 것이다.

### 참고문헌

#### 1. 원전자료

- 『계서잡록』(일사문고본·하버드대본·국립중앙도서관본·국립중앙도서관본·가람문고본·장서각본)
- 『계서야담』(천리대본·규장각본·연세대본)
- 『기문총화』(연세대4책본·동양문고갑을본·가람문고본·천리대본·국립중앙도서관본)
- 『동상기찬』(한남서림본)
- 『동야회집』(대관부립도서관본·천리대본·가람문고본·장지영본·서울대본)
- 『동패락송』(연세대본·동양문고본)
- 『선언편』(장서각본)
- 『성수총화』(연민본)
- 『쇄어』(연세대본)
- 『천예록』(천리대본·고려대본·백두산기·김영복본)
- 『청구야담』(버클리대본·동경대본·교토대본·동양문고본)
- 『총화』(천리대본)

김동욱 역, 『국역 기문총화』 1~5, 아세아문화사, 1999, 총 2170쪽.  
 \_\_\_\_\_, 『국역 청야담수』 1~3, 보고서, 2004, 총 1015쪽.  
 \_\_\_\_\_, 『국역 동패락송』 1:2, 보고서, 2012, 총 536쪽.  
 김준형, 『동상기찬』, 휴머니스트, 2012, 1~656쪽.  
 동국대학교 한국문학연구소 저, 『한국문헌설화전집』 1~10, 태학사, 1981, 총 5601쪽.  
 박용식 저, 『한국야담사화집성』 1~5, 국학자료원, 1995, 총 3899쪽.  
 정명기 저, 『한국야담자료집성』 1~24, 계명문화사, 1995, 총 14509쪽.  
 정환국 역, 『교감 역주 천예록』, 성균관대 출판부, 2005, 1~516쪽.  
 \_\_\_\_\_, 『정본 한국 야담전집』 1~10, 보고서, 2021, 총 4146쪽.

#### 2. 논저

- 강성용, 「고전텍스트의 DNA를 묻는다 : 고전텍스트의 문헌전승 계통도 분석에서 계통발생 분석방법(phylogenetical analysis)의 적용과 모색」, 『인문언어』 15(3), 국제언어인문학회, 2013, 77~122쪽.

- 강우규·김바로, 「소현성록 연작의 문체론적 고찰-컴퓨터를 활용한 계층분석을 바탕으로」, 『인문과학연구』 59, 강원대학교 인문과학연구소, 2018, 29~46쪽.
- 강혜규, 「문유채 서사의 변이 양상과 의미」, 『Journal of Korean Culture』 52, 한국어문학회국제학술포럼, 2021, 439~465쪽.
- 권기성, 「『계서잡록』 4권 소재 야담의 서사방식 - 『동패락송』과의 비교를 중심으로」, 『한국한문학회』 80, 한국한문학회, 2020, 71~105쪽.
- 권혁래, 「디지털 인문학과 고전문학의 만남」, 『열상고전연구』 50, 열상고전연구회, 2016, 9~12쪽.
- 김성문·김바로, 「디지털 분석 기법을 활용한 시조문학 연구」, 『문화와 융합』 41(5), 한국문화융합학회, 2019, 747~772쪽.
- 김영진, 「조선 후기 사대부의 야담창작과 향유의 일양상」, 『어문논집』 37, 안암어문학회, 1998, 21~45쪽.
- 김준형, 「기문총화계 야담집의 문헌학적 연구」, 고려대 석사학위논문, 1997, 1~128쪽.
- \_\_\_\_\_, 「옥소선 이야기의 변이양상과 의미」, 『한국민속학』 30, 민속학회, 1999, 265~288쪽.
- \_\_\_\_\_, 「19세기 야담 작가의 존재 양상 - 계서 이희평론」, 『민족문화사연구』 15, 민족문화사학회, 1999, 65~102쪽.
- 김현·안승준·류인태, 「데이터 기반 인문학 연구 방법의 모색 : 문중 고문서 아카이브와 디지털 인문학의 만남」, 『횡단인문학』 창간호, 숙명인문학연구소, 2018, 17~61쪽.
- 류인태, 「데이터로 읽는 17세기 재지사족의 일상 : 『지암일기(1692~1699) 데이터베이스 편찬연구」, 한국학중앙연구원 박사학위논문, 2019, 1~275쪽.
- 박경우, 「머신러닝을 활용한 고시조 연구 방법론 탐색 -운율 패턴 추출 방안을 중심으로-」, 『열상고전연구』 74, 열상고전연구회, 2021, 37~68쪽.
- 신상민 외, 「분할법에 의한 가상점을 활용한 다차원척도법」, 『응용통계연구』 28(6) 한국통계학회, 2015, 1171~1180쪽.
- 안정용, 「의학진단에 이용되는 해밍 거리의 특성 탐색」, 『한국데이터정보과학회지』 23(2) 한국데이터정보과학회, 2012, 227~234쪽.
- 우숙영 외, 「상대적 계층적 군집 방법을 이용한 마이크로어레이 자료의 군집분석」, 『한국데이터정보과학회지』 25(5), 한국데이터정보과학회, 2014, 999~1009쪽.

- 이강옥, 「야담집에서의 이경류 이야기의 전개와 그 의미」, 『한국문학논총』 61, 한국문학회, 2012, 61~93쪽.
- 이신성, 「〈옥소선 이야기〉의 전개양상과 그 의미」, 『과전김무조박사화갑기념논총』, 동간행위원회, 1988, 1~17쪽.
- 이채경, 「18세기 야담의 유사 화소 교착 및 변개 양상」, 『대동한문학』 40, 대동한문학회, 2014, 399~441쪽.
- 장진숙, 「〈옥소선 이야기〉의 전변양상과 그 서사적 의미」, 『고전문학과 교육』 16, 한국고전문학교육학회, 2008, 297~326쪽.
- 정명기, 「홍순언 이야기의 갈래와 그 의미」, 『동방학지』 45, 연세대학교 국학연구원, 1984, 301~341쪽.
- 최운호·김동건, 「군집분석 기법을 이용한 텍스트의 계통분석」, 『인문논총』 62, 서울대학교인문학연구원, 2009, 210~213쪽.
- \_\_\_\_\_, 「컴퓨터를 이용한 고전문학 디지털콘텐츠의 유사도에 따른 이본 계통 분류 연구」, 『한국정보기술학회지』 12(7), 한국정보기술학회, 2014, 101~110쪽.
- \_\_\_\_\_, 「컴퓨터 문헌 분석 기법을 활용한 〈토끼전〉 이본 연구」, 『우리문학연구』 58, 우리문학회, 2018, 123~154쪽.

Hamming, R.W, "Error detecting and error correcting codes", Bell System Technical Journal, 26, 1950, pp.147~160.

ABSTRACT

A lengthwise comparative study of different versions of Yadam  
- Based on "Ok So-seon"

Kwon, Ki-sung & Choi, Woon-ho & Kim, Dong-keon

This study uses lengthwise measurement of literary text using digital tools to classify various versions of the series "The Story of Ok So-seon." Although there have been several attempts at analyzing the development process or variation pattern of "The Story of Ok So-seon," this study applies a new methodology to compare different versions of Yadam rather than denying earlier results or revealing new aspects. Therefore, the focus was on objectively corroborating existing studies and revealing new possibilities.

First, the distances between different versions were measured and visualized as a check on earlier discussion through character unit sequence comparison and comparison of semantic types for each paragraph. As examined, the series classification of "The Story of Ok So-seon" was similar to that previous studies, and the cause of the difference was also a part that could be interpreted in the development process of Yadam Temple. However, in terms of disclosure, it was found that this study within the same series was relatively insufficient. This may be because there is no significant difference in narrative, although it was observed that objective and quantitative research of the literary summation series, which occupies a large proportion of Yadam's history, is needed.

If multiple Yadam books are examined, inevitable mistakes and inefficiencies are introduced into human processing capacity. However, until now, most studies of Yadam have been conducted following traditional humanities methods. While acknowledging their strengths, attempts to explore new possibilities should also be made. This can lead

to visualizing the actual compartments and relationships of the Yadam enjoyment pattern in the future, and the horizon of research on the different versions of Yadam can be expanded using the research methodology of a new era of digital humanities.

**Key Words**

Ok So-Seon, version study, transmission and variations, textual criticism, digital humanities

논문투고일 : 2022.03.31.

심사완료일 : 2022.05.02.

게재확정일 : 2022.05.03.