

텍스트마이닝 기법을 활용한 구비설화 전승집단의 이야기적 관심사와 그 의미

—〈한국구비문학대계〉 디지털 아카이브를 대상으로*

한유진**

〈차례〉

1. 서론
2. 텍스트 데이터 분석 과정
3. 구비설화 향유층의 이야기적 관심사의 양상
4. 젠더에 따른 전승집단의 이야기적 관심사의 차이
5. 결론

〈국문초록〉

본고는 〈한국구비문학대계〉 디지털 아카이브에 DB화된 구비설화 각편 26,542편을 대상으로 텍스트마이닝 기법을 사용하여 전체 전승집단과 젠더에 따른 전승집단의 이야기적 관심사의 양상과 그 의미를 살펴보았다. 이를 위해 KoNLPy 패키지의 Mecab 클래스를 활용하여 구비설화 26,542편의 제목에서 형태소를 분석하여 명사를 추출한 후 collection 패키지의 Counter 클래스를 사용하여 상위 빈도수를 추출하였다.

그 결과 상위 빈도수를 차지하는 30개의 명사를 통해 살펴보면 구비설화 전승집단은 ‘지명유래담’, ‘호랑이’ 설화, ‘도깨비’ 이야기, ‘가족관계의 관계성’을 보여주는 서사, ‘부·효·복·명당 등 인간이 추구하는 가치’에 대한 이야기들을 주로 구연한 것으로 확인된다.

‘지명유래담’이 가장 많이 구연된 결과는 화자들이 ‘조사할 가치가 충분한 이야기’

* 이 논문은 2022년 5월 14일 한국고전연구학회 춘계학술대회에서 발표했던 글을 수정, 보완한 것이다.

** 선문대학교 이니티움교양대학 강사

를 지명유래담으로 인식하고 있음을 보여주는 것이다. ‘호랑이’ 이야기는 향유층에게 웃음을 주고 ‘효’를 중시하는 가치를 드러냈다. ‘도깨비’ 이야기는 이야기 속 허구적 상황을 실제 사건인 것처럼 전달하여 향유층에게 ‘오싹함’을 느끼게 하는 흥미를 유발함으로써의 ‘이야기적 쾌락’을 만족시키거나 현실에서 설명되기 어려운 문제들을 해명하기 위한 수단이 되었다. ‘가족관계’를 보여주는 이야기 중에서는 특히 ‘며느리’ 소재 설화가 압도적으로 많이 구연되었는데 이는 현실에서 가장 갈등을 발생시키는 가족관계가 며느리를 둘러싼 것임을 보여주는 것이다. ‘아들’, ‘명당’, ‘효자’, ‘부자’, ‘묘’, ‘복’, ‘효부’ 등의 명사는 전승집단이 대를 이을 ‘아들을 출산’하고 ‘부자’가 되는 것에 가장 관심을 두고 있으며 다른 여러 이념보다도 특히 ‘효’를 가장 중요하게 생각하고 있음을 보여준다. 이에 대한 인물의 간절한 염원은 ‘명당’을 통해 실현되는 것으로 나타나며 ‘명당’과 ‘묘’ 소재 설화가 많이 전승되는 것도 이러한 맥락에서 해석될 수 있다.

한편 젠더에 따른 이야기적 관심사는 2,916명의 남성화자들이 구연한 15,088편, 2,004명의 여성화자들이 구연한 7,467편의 각편을 통해 살펴보았다. 남성화자들은 지명유래담을 많이 구연하였고 여성화자들은 가족 간에 발생하는 문제를 보여주는 이야기들을 즐겨 구연하였다. 이는 여성화자들의 삶이 ‘가족을 중심’으로 하는 ‘가정안’에 집중된 것을 보여주는 것으로 ‘박문수’, ‘중’, ‘정승’ 소재 이야기를 즐겨 구연하며 관심의 영역을 집박으로 확대하는 남성화자들과는 다른 것이다.

주제어 구비설화, 『한국구비문학대계』, 〈한국구비문학대계〉 디지털 아카이브, 디지털 방법론, 텍스트마이닝, 전승집단, 이야기적 관심사, 지명유래, 호랑이, 도깨비

1. 서론

본고는 『한국구비문학대계』 · 『증편 한국구비문학대계』에 수록된 구비설화가 어떠한 소재의 이야기들에 집중되어 있는지 텍스트마이닝 기법을 활용하여 그 양상을 파악해보으로써 ‘전승집단의 이야기적 관심사’를

살피는 것을 목적으로 한다. 이는 전승집단이 현실적 삶의 차원에서 주요한 가치로 여기는 것은 어떤 것인지 이들이 회구해온 보편적 욕망은 무엇 인지를 들여다볼 수 있게 해준다는 의미가 있다. 이처럼 이야기를 통해 삶에 대한 전승집단의 인식을 읽어볼 수 있는 것은 많은 사람의 흥미가 전제 되어야만 전승 가능한 구비문학 고유의 특성 때문이다.¹⁾

본고의 분석 대상은 『한국구비문학대계』와 『증편 한국구비문학대계』에 실려 있는 구비설화이다. 『한국구비문학대계』는 1979~1985년 현지 조사를 통해 수집된 구비설화를 채록하여 집대성한 자료집으로 총 82권의 단행본으로 출판되어 있으며 여기에는 설화 15,107편이 실려 있다.²⁾ 또한 『증편 한국구비문학대계』는 2008~2018년 『한국구비문학대계』 현지 조사 시 빠진 지역의 이야기를 추가로 조사한 자료집으로 이 역시 단행본으로 출간되어 있다. 이들 자료는 거의 대부분 <한국구비문학대계> 디지털 아카이브(<http://gubi.aks.ac.kr>)에 DB화되어 있는데 각편 수는 총 26,542편이다.³⁾ 이때 이들 이야기를 구연한 화자는 모두 5,178명으로 남성화자 2,916명(15,088편), 여성화자 2,004명(7,467편), 미분류된 화자 258명(3,987편)으로 나타난다.⁴⁾

1) 서대석, 『이야기의 의미와 해석』, 세창출판사, 2011, 18쪽 참고.
 2) 『한국구비문학대계』에는 이러한 설화 외에도 민요 6,187편, 무가 387편, 기타 21편이 실려 있다. (조동일, 「『한국구비문학대계』 자료 수집과 설화 분류의 원리」, 『한국구비문학대계 별책부록(I)』, 한국정신문화연구원, 1989, 1~3쪽 참고)
 3) 이 각편의 수는 『한국구비문학대계』에서 12,953편, 『증편 한국구비문학대계』에서 13,447편, 미분류된 각편 142편을 합친 것이다. 이때 <한국구비문학대계> 아카이브의 설화 수가 『한국구비문학대계』에 수록된 것보다 적은데 그 이유는 『한국구비문학대계』 82권 가운데 68권만 DB화되어 있기 때문이다. 즉 14권은 DB화되어 있지 않은 것인데 이에 해당하는 『한국구비문학대계』 책 번호를 열거하면 1-3, 1-4, 1-5, 1-6, 2-1, 2-2, 5-1, 5-3, 6-1, 6-7, 6-9, 6-10, 6-11, 6-12이다. 이들 자료는 경기도(1-), 강원도(2-), 전라북도(5-), 전라남도(6-)에서 조사된 이야기들로 조속한 DB화가 필요하다.
 4) 여기서 미분류된 258명은 <한국구비문학대계> 아카이브에서 성별 확인이 어려운

본고는 〈한국구비문학대계〉에 DB화된 구비설화 각편 26,542편을 대상으로 디지털 기법을 사용하여 이를 연행한 전승집단, 그리고 젠더에 따른 화자들의 이야기적 관심사를 파악하고 그 의미를 살피고자 하는 것이다.⁵⁾

그간 구비설화에서 어떠한 유형의 이야기들이 집중적으로 전승되었는지, 남녀 화자들은 어떠한 소재 설화에 특히 관심을 두고 구연하였는지에 대한 논의는 대부분 현장 조사 경험에 의해 가늠되거나 개별 유형 설화 연구를 통해 제시된 것으로 구비설화 전반을 대상으로 한 것이 아니었다. 이는 『한국구비문학대계』에 실린 구비설화의 존재 양상을 보여주고 있는 분류 기준과 별개로⁶⁾ 전국에서 수집된 이 방대한 자료들을 구연 빈도수 별로 재분류하는 데에 대한 어려움을 보여주는 것이다. 이러한 상황에서 『한국구비문학대계』에 수록된 14,849편의 설화를 대상으로 키워드, 지역, 유형 세 차원에서 한국설화의 네트워크 지형도를 살핀 연구는 설화의 핵심 구조를 파악하고 이야기의 지역적 특성을 찾아볼 수 있는 발판을 마련하였으며 새로운 분류 유형 작업의 가능성을 보여준 의미가 있다.⁷⁾ 이차

화자의 인원수이다. 본고에서 제시한 전체 이야기의 편수는 정확하지만 화자의 명수는 약간의 오류 가능성이 있다. 이는 〈한국구비문학대계〉 아카이브 자료가 제보자, 이름, 성별, 제목이 하나의 자료로 등록되어 있지 않음으로써 나타난 문제이다. 예컨대 김유선 화자가 구연한 설화의 링크는 김석출 화자의 〈문답가〉로 연결된다. 이에 따라 ‘제보자의 이름을 매개’하여 ‘성별과 제목’이 통합된 데이터를 구성하면 매칭되지 않는 화자가 나타나는데 그 수가 258명인 것이다.

- 5) 〈한국구비문학대계〉 아카이브에 DB화된 『한국구비문학대계』의 구비설화는 문서화된 설화보다 약 2천여 편이 적은 것이지만 그것은 특정 유형 설화가 아닌 몇 군데 지역에서 조사한 이야기들이 제외된 결과이므로 통계적 분석을 통해 구비설화 향유층 전반의 이야기적 관심사를 살피는 데에는 문제가 없을 것이라 생각한다.
- 6) 『한국구비문학대계』에 실린 구비설화는 주체와 상황에 따라 1~8유형으로 분류되어 있는데, 주체가 특이한 설화는 1~4유형, 상황이 특이한 설화는 5~8유형으로, 이러한 8개의 각 상위유형에 4개의 하위유형을 배치시켜 총 32개의 고정 상위유형으로 분류되어 있다. 이러한 분류 체계는 다음의 책에서 자세히 참고할 수 있다. 조동일, 앞의 책, 1989, 12~18쪽.

럼 다양한 차원에서 구비설화의 네트워크 지형을 살핀 이 연구는 디지털 방법론을 활용한 것이었다.

이는 방대한 양의 텍스트 양상을 구체적으로 수치화하기 위해서는 디지털 방법론이 필요함을 보여주는 것이다. 이에 따라 본고에서는 〈한국구비문학대계〉 디지털 아카이브에 DB화되어 있는 『한국구비문학대계』와 『증편 한국구비문학대계』의 구비설화 자료 분석에 텍스트마이닝 기법을 사용하고자 한다. 이는 많은 양의 비구조적인 디지털화된 텍스트를 분석하는 데에 유용한 방법으로⁸⁾ 본고에서 사용한 방법은 텍스트마이닝 기법 가운데 ‘이야기 제목’의 형태소를 분석한 후 그 중 ‘명사의 빈도수’를 조사하여 워드클라우드(word cloud)⁹⁾로 시각화한 것이다.¹⁰⁾

이때 ‘이야기의 제목’을 분석 대상으로 삼는 이유는 ‘제목’은 전체 서사를 ‘핵심적으로 가장 짧게 압축’한 것으로 화자들이 어떠한 종류의 이야기를 했는지를 보여주는 주요한 지표이기 때문이다.¹¹⁾ 이에 따라 구비설화

7) 심우장·김영원·황치욱, 「한국설화의 네트워크 지형 연구 시론」, 『구비문학연구』 37, 한국구비문학학회, 2013.

8) 김연창 외, 「텍스트 마이닝을 이용한 지능적 워드클라우드」, 『한국정보통신학회 2019년 춘계 종합학술대회 논문집』, 한국정보통신학회, 2019, 325쪽 참고.
텍스트마이닝을 활용한 텍스트 분석의 필요성과 유용성에 대해서는 이러한 기법으로 텍스트를 분석한 모든 논문에서 자세히 지적되었기 때문에 반복하여 따로 언급하지 않는다. 대표적으로 다음의 논문을 참고할 수 있다. 강우규·김바로, 「빅데이터와 고전문학 연구방법론」, 『어문논집』 78, 중앙어문학회, 2019.

9) 워드클라우드란 문서의 키워드, 개념 등을 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하는 기법이다.

10) 심우장도 개별 설화의 제목에서 키워드를 추출하는 작업을 하였는데 이때 명사뿐만 아니라 동사와 형용사도 키워드에 포함하였다. 이는 키워드들의 링크, 즉 연결관계성을 파악하기 위한 방법으로 그 결과 ‘이야기’라는 어휘가 가장 많은 링크를 가진 것으로 나타났다. 이러한 방법은 본고에서 설화의 소재를 중점적으로 살피기 위해 ‘이야기’를 키워드에서 제외한 것과는 차이가 있는 것이다. (심우장·김영원·황치욱, 앞의 글, 76~89쪽 참고)

11) 각편의 제목은 이야기를 조사하고 채록한 조사자들이 정하는 것이다. 이에 따라 개별

26,542편의 ‘제목’을 분석함으로써 화자들이 주로 전승한 이야기들을 파악하여 구비설화 향유층의 이야기적 관심사의 지형을 확인해볼 수 있는 것이다.

이처럼 구비설화 전반을 대상으로 이러한 ‘정량적 방법론’이 유용한 이유는 ‘각편의 제목에서 추출된 명사’는 ‘이야기의 주요 소재’이자 ‘서사의 내용을 대표하는 핵심어’이기 때문이다. 따라서 추출된 ‘명사의 빈도수’ 분석은 특정 이야기에 대한 ‘화자의 관심도를 수치화’하여 통계적으로 보여주는 의미가 있다.

이에 따라 본 연구에서는 『한국구비문학대계』와 『증편 구비문학대계』를 아우르고 있는 <한국구비문학대계> 아카이브의 구비설화 자료를 대상으로 전승집단의 ‘이야기적 관심사’를 살피고자 한다. 이를 위해 2장에서는 데이터 분석 과정을 서술하고 3장에서는 도출된 결과를 바탕으로 전체 구비설화 향유층의 이야기적 관심사의 양상을 4장에서는 젠더에 따른 전승집단의 이야기적 관심사의 차이를 확인하고 그 의미를 읽어볼 것이다.

2. 텍스트 데이터 분석 과정

이 장에서는 <한국구비문학대계> 아카이브 구비설화 자료에서 전체 화자, 그리고 남녀 화자집단 각각의 이야기적 관심사를 수치화하는 과정을 살피고자 한다.

조사자별로 제목을 구성하는 방법에는 다소간의 차이가 있을 수 있다. 즉 <사자생손지지> 설화 각편의 경우 ‘사자생손’으로 이야기의 유형을 바로 추정할 수 있도록 제목을 부여하기도 하지만 ‘대를 이은 은장도’와 같이 유형명에서 비껴나 있기도 한다. 사실 구비설화 각편 제목은 설화 유형이 직접 드러나지 않는 각편의 제목이 압도적으로 많다. 그렇다고 하더라도 핵심적인 이야기의 내용을 바탕으로 제목을 구성하기 때문에 ‘각편 제목’에 ‘핵심어’가 제시된다는 사실은 변하지 않으며, 설화 유형을 알고 있다면 추출된 핵심어를 통해 이야기 유형의 추정이 가능하다.

다음 제시한 [그림 1]은 이에 대한 과정을 순서대로 정리한 것으로 ‘데이터 준비→ 데이터 처리→ 데이터 시각화’로 요약할 수 있다.



[그림 1] <한국구비문학대계> 구비설화 각편 제목의 명사 빈도수 시각화 과정

먼저 ‘데이터 준비’ 과정이다.¹²⁾ <한국구비문학대계> 구비설화 제목의 명사 빈도수를 통계화하기 위해서는 일차적으로 ‘전체 구비설화 자료의 제목’이 필요하다. 이를 위해 아래 [그림 2]와 같이 <한국구비문학대계> 아카이브 ‘유형분류별→설화’ 항목에 속한 자료에서 각편 26,542편의 아카이브를 긁어온 후 [유형별 데이터]를 생성했다. 이때 이러한 [유형별 데이터] 자료에는 성별에 대한 정보가 없기 때문에 남녀 화자집단에 따른 이야기의 빈도수를 살피기 위해서는 추가적인 정보가 필요하다. 이에 따라 <한국구비문학대계> 아카이브 ‘제보자별’ 항목 9,742건을 스크래핑하고 [제보자별 데이터]를 만들었다.

12) 데이터 전처리는 MATLAB 프로그램을 활용하였다.

| 번호 | 유형 | 제목 | 싱크 | 제보자 | 연도 | 소재지 |
|----|----|---------------|----|-----|------|------------------------------------|
| 1 | 설화 | "전라도 개동새"의 유래 | O | 이민희 | 1982 | 충남 부여군 은산면 |
| 2 | 설화 | '곳'의 유래 | O | 강승호 | 2016 | 경기 성남시 수정구 삼성대로 215번길 7(신흥동) 성남문화원 |

[그림 2] [유형별 데이터] 아카이브

| 번호 | 이름 | 성별 | 생년(연월) | 지역 | 제공자료 |
|----|-----|----|-------------|--------------|--|
| 1 | 김병철 | 남 | 1943년생 (정미) | 경기 고양시 일산... | 설화 0, 민요 2, 무가 0, 현대구전설화 0, 근현대구전민요 0, 무령 0, 기타 0, 전체 2 |
| 2 | 가신행 | 여 | 1921 | 충남 서산시 부... | 설화 0, 민요 3, 무가 0, 현대구전설화 3, 근현대구전민요 0, 무령 0, 기타 0, 전체 14 |

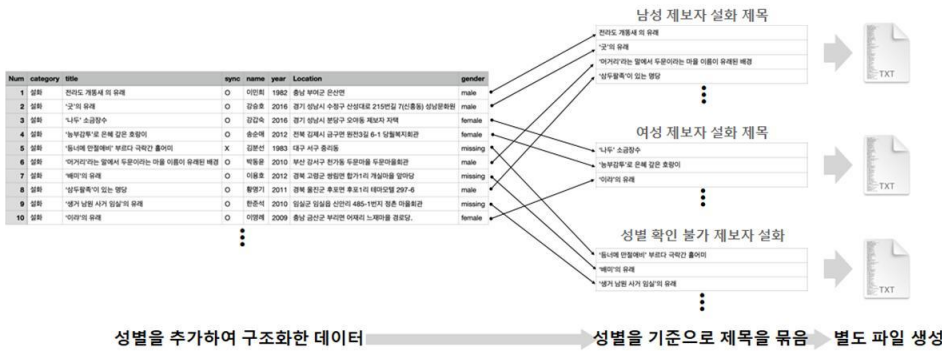
[그림 3] [제보자별 데이터] 아카이브

그런 다음 ‘데이터를 구조화’하는 작업을 한다. 즉 [유형별 데이터]와 [제보자별 데이터]에서 ‘제보자의 이름을 매개’하여 ‘성별과 제목’이 통합된 ‘테이블 데이터’를 아래 [그림 4]와 같이 구성했다. 이는 <한국구비문학대계> 아카이브에서 [유형별 데이터]에는 성별이 제시되어 있지 않고 [제보자별 데이터]에서는 제목을 확인할 수 없기 때문이다.

| Num | category | title | sync | name | year | Location | gender |
|-----|----------|---------------------------------|------|------|------|------------------------------------|---------|
| 1 | 설화 | 전라도 개동새 의 유래 | O | 이민희 | 1982 | 충남 부여군 은산면 | male |
| 2 | 설화 | '곳'의 유래 | O | 강승호 | 2016 | 경기 성남시 수정구 삼성대로 215번길 7(신흥동) 성남문화원 | male |
| 3 | 설화 | '나두' 소금장수 | O | 강갑숙 | 2016 | 경기 성남시 분당구 오이동 제보자 자택 | female |
| 4 | 설화 | '농부김투'로 인해 같은 호랑이 | O | 송순애 | 2012 | 전북 김제시 금구면 원전3길 6-1 당월복지회관 | female |
| 5 | 설화 | '동녀에 만짐에비' 부르다 극락간 흠어미 | X | 김분선 | 1983 | 대구 서구 중리동 | missing |
| 6 | 설화 | '머거리'라는 말에서 두문이라는 마을 이름이 유래된 배경 | O | 박동운 | 2010 | 부산 강서구 천가동 두문마을 두문마을회관 | male |
| 7 | 설화 | '배미'의 유래 | O | 이용호 | 2012 | 경북 고령군 병림면 합가1리 개실마을 앞마당 | missing |
| 8 | 설화 | '삼두팔죽'이 있는 명당 | O | 황영기 | 2011 | 경북 울진군 후포면 후포1리 테미모텔 297-6 | male |

[그림 4] 테이블 데이터

이것을 바탕으로 말뭉치를 만든다. 이것은 다음 [그림 5]와 같은 방법으로 젠더에 따라 제목을 분류하여 모은 텍스트 파일인데 남녀 화자별로 이야기를 분류하기 위한 것이다.

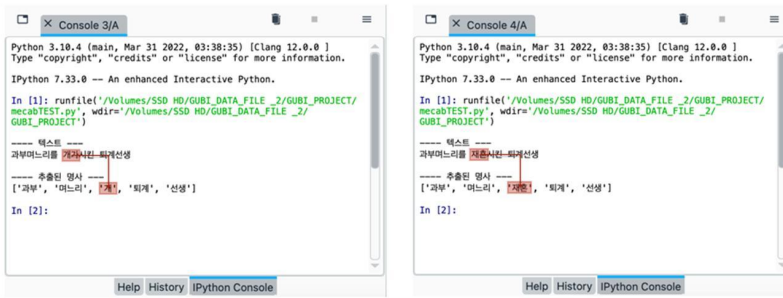


[그림 5] 젠더에 따른 제목 분류 및 말뭉치 생성

다음은 ‘데이터 처리’ 과정이다. 우선 파이썬(Python) 언어를 기반으로 KoNLpy 패키지의 Mecab 클래스를 활용하여 형태소를 분석했다.¹³⁾ 이렇게 분석된 형태소는 파이썬의 collection 패키지의 Counter 클래스를 활용하여 상위 빈도수를 추출하였다. 이때 KoNLpy를 이용한 형태소 분석 과정에서 발생하는 다음 [그림 6(a)]과 같은 오류들에 대해서는 [그림 6(b)]와 같이 이를 수정하는 작업을 하였다. 예컨대 ‘과부며느리를 개가시킨 퇴계선생’의 제목에서 형태소를 분석하면 ‘과부’, ‘며느리’, ‘개’, ‘퇴계’, ‘선생’이 명사로 추출된다. 이처럼 ‘개가’가 아닌 ‘개’가 나타나는 이유는 ‘개가’를 명사로 처리하지 못하기 때문인 것으로¹⁴⁾ ‘개가’를 ‘재혼’으로 바꾸어서 프로그램을 실행하는 방법을 사용하여 그 오류를 개선하였다. 이렇게 수정된 데이터 결과를 바탕으로 ‘명사’ 및 ‘명사의 빈도수’를 추출하였다.

13) Macab 클래스를 사용한 이유는 KoNLpy 패키지의 형태소 분석기 가운데 실행 속도가 가장 빠르기 때문이다. 다양한 형태소 분석기의 성능에 대해서는 다음의 논문을 참고할 수 있다. 원혜영·이현영·강승식, 「대규모 텍스트 분석을 위한 한국어 형태소 분석기의 실행 성능 비교」, 『2020년 한국컴퓨터종합학술대회 논문집』, 한국정보과학회, 2020.

14) 이러한 이유는 ‘개가’를 명사 ‘개’와 조사 ‘가’로 분류하기 때문인 것으로 보인다.



(a) (b)

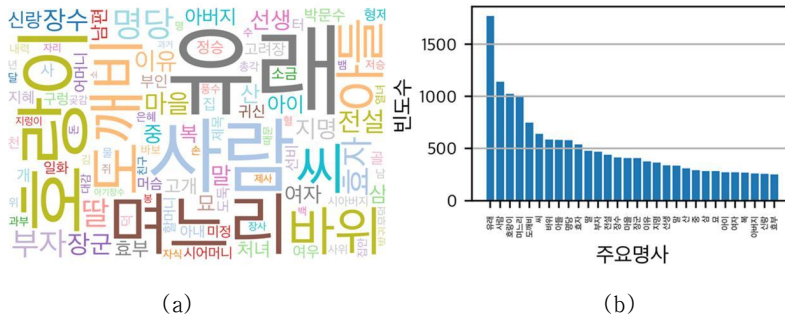
[그림 6] (a)분류 오류의 예시, (b)오류 수정의 예시

이러한 과정을 거친 결과는 히스토그램과 워드클라우드로 시각화하였다. 이는 다음 3장과 4장에서 자세히 분석하고자 한다.

3. 구비설화 향유층의 이야기적 관심사의 양상

이 장에서는 2장에서 분석한 데이터인 <한국구비문학대계> 아카이브 ‘구비설화 전체 자료의 각 제목’을 구성하는 ‘명사의 빈도수를 시각화’하고 그 결과를 바탕으로 ‘전승집단이 어떠한 소재의 이야기를 집중적으로 구연’했는지를 파악하여 그 의미를 살피고자 한다.

<한국구비문학대계> 아카이브의 구비설화는 총 26,542편으로 각 이야기 제목에서 명사 빈도수를 추출하면 아래 [그림 7]과 같은 형태로 나타난다. 다음 제시한 [그림 7(a)]와 [그림 7(b)]는 각각 명사의 빈도수를 보여주는 워드클라우드와 히스토그램으로, 워드클라우드에는 상위 100개, 히스토그램의 경우는 30개의 명사를 제시하였다.



(a) (b)
 [그림 7] 구비설화 26,542편 제목의 명사 빈도수에 대한
 (a)워드클라우드 (b)히스토그램

히스토그램에 제시한 빈도수 상위 30개 명사의 수치를 표로 제시하면
 다음 [표 1]과 같다.¹⁵⁾

| 순위 | 단어 | 빈도수 | 순위 | 단어 | 빈도수 | 순위 | 단어 | 빈도수 |
|----|-----|-------|----|----|-----|----|-----|-----|
| 1 | 유래 | 1,771 | 11 | 딸 | 477 | 21 | 산 | 310 |
| 2 | 사람 | 1,139 | 12 | 부자 | 469 | 22 | 증 | 293 |
| 3 | 호랑이 | 1,023 | 13 | 전설 | 441 | 23 | 삼 | 284 |
| 4 | 며느리 | 995 | 14 | 장수 | 414 | 24 | 묘 | 284 |
| 5 | 도깨비 | 784 | 15 | 마을 | 410 | 25 | 아이 | 272 |
| 6 | 씨 | 640 | 16 | 장군 | 409 | 26 | 여자 | 272 |
| 7 | 바위 | 586 | 17 | 이유 | 378 | 27 | 복 | 268 |
| 8 | 아들 | 582 | 18 | 지명 | 366 | 28 | 아버지 | 261 |
| 9 | 명당 | 579 | 19 | 선생 | 340 | 29 | 신랑 | 258 |
| 10 | 효자 | 540 | 20 | 말 | 337 | 30 | 효부 | 253 |

[표 1] 구비설화 26,542편 제목의 명사 빈도수

15) 이때 제목에서 나타나는 ‘이야기’는 불용어(不用語) 처리를 하였으며 이는 그 결과이다. 실제 가장 높은 빈도수를 보이는 명사는 ‘이야기’로 나타나는데 총 2,007건이다. ‘이야기’라는 어휘는 서사의 내용과는 무관한 것이므로 본 연구에서는 불용어로 처리한 후 워드클라우드와 히스토그램을 제시한 것이다. 이에 대해서는 4장 자료들의 경우도 동일하게 적용하였다.

위 [표 1]에 제시된 명사의 빈도수를 바탕으로 구연된 이야기의 범주를 설화 분류 유형에 따라 정리하면 ‘신화적 내용’에 대한 이야기, ‘동물담’, ‘괴기담(怪奇譚)’, ‘인간담’, ‘신앙담’으로 간추려진다.¹⁶⁾ 이때 화자들은 신화적 내용으로 분류된 유형 중 ‘지명유래담’, 동물담 가운데 ‘호랑이’ 설화, 괴기담에서 ‘도깨비’ 이야기, 인간담에 해당하는 ‘고부(姑婦)·부모와 자식·부부 등 가족관계의 관계성’을 보여주는 서사, 신앙담에 속하는 ‘부(富)·효(孝)·복·명당 등 인간이 추구하는 가치’에 대한 이야기들을 주로 구연한 것이 확인된다. 따라서 구비설화에서 많은 비중을 차지하는 이들 다섯 유형의 설화들을 차례로 살피고 화자들이 이 유형의 설화들을 집중적으로 구연한 이유가 무엇인지를 탐색해보고자 한다. 이는 구비설화 향유층의 이야기적 관심사의 양상과 그 의미를 보여주는 의미가 있다.

먼저 ‘지명유래담’은 마을의 지형이나 특정 지명이 형성된 내력에 대한 이야기로 구연 현장에서 화자들이 많이 구연한 설화 유형 중 하나이다.¹⁷⁾ 위 [표 1]에서 제일 높은 빈도수를 보여주는 명사가 ‘유래’(1위·1,771건)라

16) 설화 분류 유형은 이야기의 ‘내용’을 중심으로 한국설화를 9개로 나눈 장덕순의 분류를 수용한 것이다. 이러한 분류에 따라 이야기를 가늠해본 것은 구연 양상을 살피는데 참고하고자 한 것으로 분류유형별 빈도수를 분석하고자 함은 아니다. 장덕순이 분류한 9개의 유형은 ‘신화적 내용’, ‘동물담’, ‘인생담’, ‘인간담’, ‘신앙담’, ‘영웅담’, ‘괴기담’, ‘소화’, ‘형식담’으로 이를 간략히 설명하면 다음과 같다. 즉 ‘신화적 내용’은 자연현상 및 인문현상의 기원과 유래를 설명한 이야기이고, 동물을 대상으로 한 모든 설화는 ‘동물담’, 인간의 통과의례를 다룬 ‘인생담’, 인물들 간 맺는 관계에서 발생하는 내용으로 구성된 ‘인간담’, 관념적인 신앙 가치를 드러내는 ‘신앙담’, 영웅적 행위를 하는 특수한 인물에 대한 서사인 ‘영웅담’, 비현실적인 세계를 보여주는 ‘괴기담’, 웃음을 주는 ‘소화’, 이외 기타 이야기가 ‘형식담’인 것이다. (장덕순, 『한국설화문학연구』, 박이정, 1995, 10~14쪽 참고)

17) 이때 ‘지명유래담’은 ‘지명 전설’도 아우르는 것이다. 서사의 설화적 구조의 유무에 따라 지명유래담과 지명 전설을 나누기도 하지만 본고에서는 이들을 구분하지 않고 ‘지명유래담’으로 통칭하여 다루기로 한다. 지명유래담과 지명 전설의 구체적인 차이에 대해서는 다음 논문을 참고할 수 있다. 임재해, 『설화의 미디어 기능과 지명전설의 인문지리 정보』, 『한민족어문학』 69, 한민족어문학회, 2015, 56쪽.

는 점은 가장 높은 비중으로 전승된 이야기가 ‘유래담’임을 추정케 하는데,¹⁸⁾ 이 가운데에서도 특히 ‘지명유래담’이 많이 연행되었다. 이는 상위 빈도수에 위치한 명사인 ‘바위’, ‘마을’, ‘지명’, ‘산’을 통해서도 이미 짐작된 것이지만 유래담에 해당하는 각편 제목의 키워드를 추출함으로써 분명히 확인된다. 즉 ‘유래’를 포함한 제목을 가진 설화 1,794편을 대상으로 명사 빈도수를 추출하면 그 결과는 지명(318건), 마을(108건), 바위(86건), 고려장(75건), 골(70건), 고개(58건), 산(51건) 등의 순서로 나타난다.

이처럼 화자들이 ‘지명유래담’을 많이 구연한 이유는 조사 당시 형성된 이야기관의 영향력을 배제할 수 없는 것인데 이는 이야기관의 구조에 대한 논의와 별개로 지명유래담에 대한 전승 인식이 드러난다는 점에서 주목된다. 즉 이들 설화가 ‘이야기 조사’를 위해 인위적으로 형성된 이야기관에서 연행되었다는 점으로 인해 오히려 화자들이 ‘기록에 남길 만한 이야기’로 무엇을 선택했는지가 드러나는 것이다. 실제로 조사자들이 각 지역의 마을 회관을 찾아가 이야기를 청했을 때 화자들이 가장 먼저 들려주는 설화가 바로 그 지역의 산이나 바위, 그 마을 지명의 유래담인 경우가 많다.¹⁹⁾ 이는 화자들이 ‘조사할 가치가 충분한 이야기’의 가장 최우선의 자리에 그 지역 설화를 놓고 있는 것으로 전승의 맥락에서 지명유래담의 권위를 드러낸다.²⁰⁾

18) 유래담은 지명에 대한 것 외에도 <김녕 김씨 유래> 와 같은 씨족유래담, <북두칠성의 유래>, <고시래의 유래> 처럼 특정 대상이나 풍속의 기원에 대한 이야기 등 다양한 유형이 있다.

19) 이는 명확히 말하면 특히 남성화자들의 특징이다. 즉 남성화자들은 조사자들이 이야기를 청할 때 가장 먼저 꺼내는 이야기가 지명유래담이지만 여성화자는 이와 다르다. 이에 대해서는 4장에서 자세히 논의하기로 한다.

20) 실제로 조사자가 이야기를 청하였을 때 화자가 지명유래담을 모르면 ‘이야기를 모른다’며 구연을 거부하는 사례가 많다. 이 경우 조사자는 호랑이, 며느리, 박문수 등 구체적인 이야기 소재를 제시하면서 다시 이야기를 요청하는데 지명유래담은 ‘이야기’를 요청하면 특별한 설명 없이도 구연한다.

그 다음으로 ‘호랑이’ 소재 설화는 [표 1]의 어휘 빈도수(3위·1,023건)에서도 짐작되듯이 화자들이 즐겨 향유하는 이야기로 실제 조사 현장에서도 대다수가 짧게나마 한두 편 정도는 구연 가능할 정도로 향유층에게 매우 친숙한 설화이다.²¹⁾ 호랑이 이야기는 다양한 유형이 전승되는데 그중에서도 많은 비중을 차지하는 설화는 ‘호랑이와 꽃감’, ‘효에 감응하는 호랑이’이다.²²⁾

먼저 ‘호랑이와 꽃감’ 설화는 호랑이가 꽃감이 자신보다 무서운 존재인 줄 알고 도망치는 이야기로 전승 현장에서 소화로 향유된다. 여기서 대적할 상대가 없는 맹수가 고작 ‘꽃감’을 두려워하였다는 점이 향유층들의 웃음을 유발한 것인데 한편으로는 ‘위협적인 호랑이의 동물성’을 무너뜨린 것이 호랑이의 ‘무지’임을 보여줌으로써 ‘앓의 중요성’을 역설하는 이야기이기도 하다.²³⁾ 한편 ‘효에 감응하는 호랑이’ 이야기는 호랑이가 시묘 사는 효자를 지켜주거나 홍시를 구하는 등의 서사로 효 설화의 한 유형으로 ‘효’를 중시하는 전승집단의 가치를 드러낸다.

이처럼 ‘호랑이’ 설화는 허구적 상상력이 작동하는 이야기의 세계를 잘 보여주는데 이는 ‘도깨비’ 설화에서도 마찬가지이다. ‘도깨비’는 부귀를 가져다주는 우리 민족의 토착 신격으로 조화에 능한 존재이다.²⁴⁾ 도깨비 이

21) 호랑이 설화는 전국에 걸쳐 광범위하고 활발하게 전승되는데 이 자체가 구비설화의 특징이기도 하다. 고전서사 가운데 호랑이 소재 이야기는 고소설 〈호질〉을 제외하고는 찾아보기 어렵다. (장덕순(1995), 앞의 책, 98쪽 참고)

22) 이는 ‘호랑이’라는 어휘가 포함된 각편 1,016편의 제목에 나타난 명사 빈도수를 추출한 결과에 따른 것이다. 많은 빈도수를 차지한 어휘는 꽃감 114건, 이야기 96건, 사람 72건, 효자 55건, 할머니 37건, 여자 27건 등의 순으로 나타난다.

23) 이에 대해서는 서대석도 다음과 같이 지적하였다. “모르기 때문에 사태를 정확히 파악하지 못하여 실패한다는 것으로 앓의 중요성을 깨닫게 하는 부분이다. (중략) 이 이야기에는 아는 자와 모르는 자의 유리함과 불리함, 상대를 다루는 수법으로 당근과 채찍의 효용성 문제가 함축되어 있다.” (서대석(2011), 앞의 책, 212쪽)

24) 김경혜, 「〈도깨비를 이용해 부자되기〉 설화에 나타난 신화성 연구」, 『인문논총』 43, 경남대학교 인문과학연구소, 2017, 200쪽.

야기는 다양한 유형이 있지만 화자들이 즐겨 구연하는 대표적인 설화로는 ‘도깨비에 홀린 사람’, ‘도깨비 덕에 부자된 사람’ 등이 있다.²⁵⁾

‘도깨비에 홀린 사람’은 도깨비 유형 가운데 양적으로 가장 많은 비중을 차지하는 이야기로 조화에 능한 도깨비의 속성에 따른 것이다.²⁶⁾ 화자들은 대부분 서사에서 도깨비에 홀려 어이없는 행동을 했거나 아프거나 죽는 이야기를 실제 사건인 것처럼 구연하는 모습을 보인다. 즉 화자는 자신의 주변 사람이 겪은 일이거나 주위 사람들에게 들은 내용을 전달하는 방식으로 구연하는 것이다. 화자들의 이러한 구연 전략은 이야기 속 허구의 상황을 실제 사건인 것처럼 전달하여 ‘오싹함’을 느끼게 하는 흥미를 유발함으로써 ‘이야기적 쾌락’을 만족시키기 위한 것으로 볼 수 있다. 또 한편으로는 이야기를 통해 현실을 이해하고자 하는 목적에서 전승된 이 유형 특유의 구연 전략으로도 읽어볼 수 있다. 즉 전승집단은 현실에서는 설명되기 어려운 문제들, 특히 ‘원인 모를 병이나 죽음’과 같은 두려움을 ‘도깨비에게 홀림’이라는 허구적 상상력을 통해 이해하고자 한 것이다.

한편 ‘도깨비 덕에 부자된 사람’ 유형은 풍요를 관장하는 도깨비의 특성이 나타나는 서사로 도깨비로 인해 부를 얻게 된 사람의 이야기를 다루고 있다. 이는 ‘부’에 대한 전승집단의 바람을 보여주는 동시에 그것이 도깨비를 만나야만 가능하다는 점에서 부를 얻는 것이 얼마나 어려운 것인지에 대한 전승집단의 현실 인식을 드러낸다.

25) 이는 ‘도깨비’ 어휘가 포함된 각편 819편의 제목에 나타난 명사 빈도수를 추출한 결과에 따른 것으로 사람 131건, 씨름 68건, 도깨비불 64건, 빗자루 45건, 부자 41건, 방망이 24건 등의 순으로 나타난다.

26) 도깨비 설화에서 추출된 명사 중 ‘사람’(131건), ‘씨름’(68건), ‘도깨비불’(64건)의 어휘가 나타나는 이야기는 도깨비에 홀린 서사로 이 유형이 도깨비 설화 중 가장 많은 비중을 차지함이 확인된다. 여기서 ‘사람’은 다른 도깨비 유형의 제목에서도 추출될 수 있는 어휘이지만 ‘도깨비에 홀린 사람’ 서사에서 가장 많이 보인다. 또한 이 어휘를 제외하고 ‘씨름’, ‘도깨비불’만을 근거로 하더라도 ‘도깨비에 홀린 서사’가 제일 많이 구연되었다는 사실은 변하지 않는다.

전승집단은 이처럼 허구적이고 비현실적인 세계의 이야기들을 통해 현실에서는 설명되지 않는 미지의 세계를 나름의 방식으로 이해하고자 하였고 ‘효’와 같이 삶에서 중요한 가치를 다시금 되새겼으며 돌연 행운을 얻게 된 사람의 서사에서 대리만족을 느낀 것이다. 이처럼 이야기는 허구적인 상상력의 세계를 통해 현실 문제를 조율하는 기능을 하였는데 일상에서 발생하는 문제들을 고스란히 보여주는 이야기들 또한 향유층들은 매우 많은 관심을 가지고 전승하였다.

이에 해당하는 이야기가 바로 ‘가족관계에서 발생하는 문제’를 다룬 설화이다. ‘가족관계 설화’는 위 [표 1]에서 제시된 명사 가운데 ‘며느리’(995건), ‘아들’(582건), ‘효자’(540건), ‘딸’(477건), ‘아버지’(261건), ‘신랑’(258건), ‘효부’(253건)를 통해 ‘고부 관계’, ‘부모 자식 관계’, ‘부부 관계’에 대한 다양한 이야기들이 전승되고 있음이 확인된다.

이때 가장 많이 구연되는 유형은 ‘며느리’ 소재 설화로 이는 현실에서 가장 갈등을 발생시키는 가족관계가 며느리를 둘러싼 것임을 보여주는 것이다. 며느리는 시집오면 시가라는 낯선 곳에 적응해야 하며 시집 식구들도 며느리를 새로운 가족 구성원으로 받아들여야 한다. 그 과정에서 서로 다른 집안 문화에서 살아왔던 이들은 예기치 못한 많은 갈등으로 인해 각기 나름의 입장에서 고충을 겪는다. ‘며느리’ 설화는 이러한 현실 문제에 대한 전승집단의 절대적 공감을 바탕으로 화자들에게 즐겨 구연되며 전승되어 온 것으로 보인다.

마지막으로 [표 1]에서 ‘아들’, ‘명당’, ‘효자’, ‘부자’, ‘묘’, ‘복’, ‘효부’ 등의 명사는 전승집단의 욕망, 이들이 중요하게 생각하는 가치를 보여주는 이야기들을 가리킨다. 즉 전승집단은 대를 이을 ‘아들을 출산’하고 ‘부자’가 되는 것에 가장 관심을 두고 있으며 다른 여러 이념보다도 특히 ‘효’를 가장 중요하게 생각하고 있음이 드러난다. 특히 이들이 바라는 욕망은 한국인의 보편적 욕망의 모습이라는 점에서 주목된다.

한국인은 수(壽), 부(富), 귀(貴), 다남(多男)을 중요하게 생각한다고 알려져 있는데,²⁷⁾ 이는 이야기를 통해서도 나타난다. 즉 ‘부유하고[富] 높은 벼슬에 올라[貴] 오래 살면서[壽] 자식을 많이 두는 것[多男]’은 구비설화에서 등장인물이 누릴 수 있는 가장 큰 복으로 형상화된다. 구비설화에서 이에 대한 인물의 간절한 염원은 ‘명당’을 통해 실현되는 것으로 나타나며 이는 전승집단의 소망이기도 한 것이다. ‘명당’과 ‘묘’ 소재 설화가 많이 전승되는 것도 이러한 맥락에서 해석될 수 있다.

4. 젠더에 따른 전승집단의 이야기적 관심사의 차이

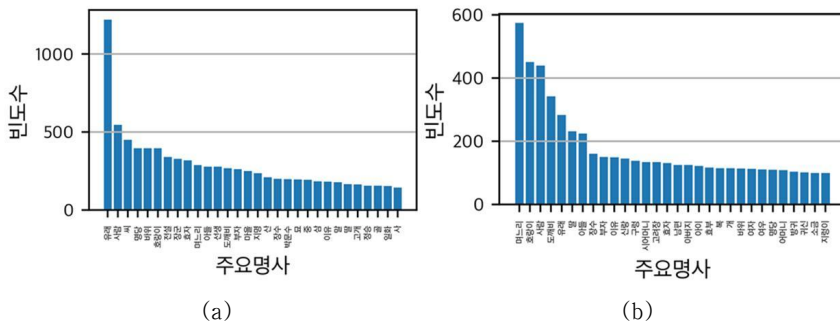
이 장에서는 〈한국구비문학대계〉 아카이브 자료를 남녀 화자들이 구연한 설화로 각각 나누어 분석한 데이터를 바탕으로 ‘젠더에 따른 향유층의 이야기적 관심사의 차이’를 살피고자 한다. 분석 대상 설화는 2,916명의 남성화자들이 구연한 15,088편, 2,004명의 여성화자들이 구연한 7,467편의 이야기이다.²⁸⁾ 〈한국구비문학대계〉 아카이브 설화 자료에서 화자의 젠더에 따른 전승 양상은 상이한 형태로 나타나는데, 이를 워드클라우드와 히스토그램으로 시각화하면 각각 다음 [그림 8], [그림 9]와 같다.

27) 동양문화의 오복(五福)에 대하여 『서경(書經)』에서는 ‘첫째는 장수하는 것이요, 둘째는 부유한 것이며, 셋째는 건강하고 평안한 것이며, 넷째는 덕을 좋아하는 것이며, 다섯째는 천수를 다하는 것으로 제시하고 있다. 하지만 한국인의 복에 관한 연구에 따르면, 우리 민족은 ‘수(壽), 부(富), 귀(貴), 다남(多男)’을 중요하게 생각한다고 한다(정재민, 『설화에 나타난 한국인의 행복관』, 『한국인의 삶과 구비문학』, 집문당, 2002, 104~106쪽 참고).

28) 여기에서 포함되지 않은 이야기는 화자 258명이 구연한 3,987편이다. 이처럼 상당히 많은 수의 설화가 젠더별로 분류한 데이터에는 빠져 있지만 이들 이야기 제목 명사의 빈도수를 분석한 결과 전체 설화를 대상으로 한 것과 그 양상이 유사하다. 따라서 미분류된 자료는 남녀 화자별 자료를 분석한 결과에는 영향을 주지 않는 것으로 추정할 수 있다.



[그림 8] 워드클라우드 (a)남성 화자집단 (b)여성 화자집단



[그림 9] 히스토그램 (a)남성 화자집단 (b)여성 화자집단

위 [그림 8]과 [그림 9]에서 가장 확연히 눈에 띄는 것은 ‘지명유래담’은 집중적으로 남성화자들에 의해 구연되었다는 것이다. 특히 남성화자들의 이야기 빈도수를 보여주는 [그림 9(a)]의 히스토그램을 보면 ‘지명유래담’은 압도적인 비율을 차지한다. 즉 남성화자들은 그 어떠한 이야기들보다도 지명유래담을 많이 구연한 것이다. 이를 보다 상세히 살피기 위해 상위 30위의 빈도수를 제시하면 다음의 [표 2], [표 3]과 같다.

여기에서 [표 2]는 남성 화자집단, [표 3]은 여성 화자집단에 해당한다.

| 순위 | 단어 | 빈도수 | 순위 | 단어 | 빈도수 | 순위 | 단어 | 빈도수 |
|----|-----|-------|----|-----|-----|----|----|-----|
| 1 | 유래 | 1,221 | 11 | 아들 | 278 | 21 | 중 | 193 |
| 2 | 사람 | 545 | | 선생 | 278 | 22 | 삼 | 184 |
| 3 | 씨 | 449 | 13 | 도깨비 | 268 | 23 | 이유 | 182 |
| 4 | 명당 | 395 | 14 | 부자 | 261 | 24 | 딸 | 178 |
| | 바위 | 395 | 15 | 마을 | 250 | 25 | 딸 | 166 |
| | 호랑이 | 395 | 16 | 지명 | 236 | 26 | 고개 | 164 |
| 7 | 전설 | 339 | 17 | 산 | 209 | 27 | 정승 | 156 |
| 8 | 장군 | 327 | 18 | 장수 | 199 | 28 | 골 | 155 |
| 9 | 효자 | 317 | 19 | 박문수 | 197 | 29 | 일화 | 154 |
| 10 | 머느리 | 287 | 20 | 묘 | 196 | 30 | 사 | 144 |

[표 2] 남성 화자집단의 구비설화 15,008편 제목의 명사 빈도수

| 순위 | 단어 | 빈도수 | 순위 | 단어 | 빈도수 | 순위 | 단어 | 빈도수 |
|----|-----|-----|----|------|-----|----|-----|-----|
| 1 | 머느리 | 574 | 11 | 신랑 | 145 | 22 | 개 | 114 |
| 2 | 호랑이 | 451 | 12 | 구렁 | 138 | | 23 | 바위 |
| 3 | 사람 | 439 | 13 | 시어머니 | 134 | 24 | 여자 | 112 |
| 4 | 도깨비 | 342 | | 고려장 | 134 | 25 | 여우 | 110 |
| 5 | 유래 | 283 | 15 | 효자 | 131 | 26 | 명당 | 109 |
| 6 | 딸 | 231 | 16 | 남편 | 125 | 27 | 어머니 | 108 |
| 7 | 아들 | 224 | | 아버지 | 125 | 28 | 방귀 | 103 |
| 8 | 장수 | 160 | 18 | 아이 | 122 | 29 | 귀신 | 101 |
| 9 | 부자 | 150 | 19 | 효부 | 117 | 29 | 소금 | 99 |
| 10 | 이유 | 149 | 20 | 복 | 114 | | 지렁이 | 99 |

[표 3] 여성 화자집단의 구비설화 7,467편 제목의 명사 빈도수

남성화자들이 구연한 설화에서는 ‘유래’(1위·1,221건), ‘바위’(4위·395건), ‘마을’(15위·250건), ‘지명’(16위·236건), ‘산’(17위·209건), ‘고개’(26위·164건), ‘골’(28위·155건) 등 지명유래담에 해당하는 명사가 제목의 상위 빈도수에 추출된 것이 [표 2]를 통해 확인된다. 반면 여성화자들의 이야기에서는 [표 3]에 제시되어 있듯이 ‘유래’가 5위(283건)이며 유래담과 연관된 단어는 ‘바위’(22위·113건)뿐이다.

이는 남성화자들은 지명유래담을 많이 구연하였지만 여성화자들은 그렇지 않음을 보여주는 것이다. 이러한 젠더에 따른 구연 경향의 차이는 화자들이 보유한 이야기의 목록 및 관심사가 다른 것으로부터 그 원인을 찾을 수 있다.

이야기 조사 시 그 지역 관련 이야기를 청하면 대부분의 남성화자들은 즉시 이에 응하지만 여성화자들은 잘 모른다고 하거나 시큰둥하게 반응하며 다른 이야기를 들려주는 경우가 많다. 이처럼 같은 마을에 살면서도 그 마을과 관련한 이야기를 남성들은 알지만 여성들은 잘 모르는 이유는 화자들의 이야기 목록이 대부분 유년 시절에 형성된 것이기 때문이다.

즉 화자들이 구연한 이야기를 어떻게 알게 된 것이냐고 그 출처를 물으면 대개 어렸을 때 들었던 이야기라고 답한다. 이야기 조사 시 남성화자들은 그 지역에서 나고 자란 토박이들이지만 여성화자들은 다른 지역에서 시집온 경우가 대부분이다. 이에 따라 여성화자들의 이야기 목록에는 현재 살고 있는 마을과 관련된 유래담은 없는 것이다.²⁹⁾ 이처럼 남성화자들은 조사 시 지명유래담을 매우 중요하게 여기며 흥미롭게 구연하지만³⁰⁾ 여성화자들에게서는 이러한 인식을 찾아보기 어렵다.

대신 여성화자들의 이야기에서는 ‘가족관계’를 나타내는 명사들이 많이 추출된 것이 특징이다. 이에 해당하는 어휘는 ‘며느리’(1위·574건), ‘딸’(6위·231건), ‘아들’(7위·224건), ‘신랑’(11위·145건), ‘시어머니’(13위·134

29) 이야기 현장에서 여성들은 실제로 ‘다른 지역에서 시집왔기 때문에 여기(지금 살고 있는 곳) 이야기는 모른다’고 말한다.

30) 김영희는 이를 남성화자들의 특성으로 지적하며 다음과 같이 논의하였다. “남성 연행자들에게 현지조사는 ‘공식적인 일’의 성격을 띠게 되는데 이때 이 공식적인 일의 내용은 ‘자신이 살고 있는 마을 공동체의 역사를 드러내고 이를 선전하는 일’로 해석된다. (중략) 이들의 연행 동기는 마을 공동체에 대한 자부심과 마을 공동체 내 전통에 대한 수호 의지, 교육적 차원에서의 사명감 등으로 구성되어 있다.” (김영희, 『마을 지형 및 지명 유래담의 공동체 구성력 탐구』, 『비교민속학』 46, 비교민속학회, 2011, 609쪽)

건), ‘고려장’(13위·134건), ‘효자’(15위·131건), ‘남편’(16위·125건), ‘아버지’(16위·125건), ‘효부’(19위·117건), ‘어머니’(26위·108건), ‘방귀’(27위·103건), ‘지렁이’(29위·99건)로 [표 3]에서 제시한 30개의 명사 가운데 13개인 거의 절반에 가까운 비율이다.³¹⁾ 이는 [표 2]에서 남성화자들의 것을 동일하게 살펴보았을 때 ‘효자’(9위·317건), ‘며느리’(10위·287건), ‘아들’(11위·278건), ‘딸’(25위·166건) 등 4개만 나타나는 것과는 분명한 차이를 보이는 것이다.

이처럼 여성화자들은 가족 간에 발생하는 문제를 보여주는 가족 설화를 즐겨 구연한 것을 알 수 있는데 이는 이들의 주요 관심이 ‘가정’에 있음을 보여준다. 특히 여성화자들의 이야기 목록에서만 보이는 ‘딸’·‘아버지’, ‘시어머니’·‘효부’, ‘신랑’·‘남편’은 핵심 서사가 부모, 고부, 부부 관계에서 발생한 사건임을 짐작케 한다. 이처럼 여성화자들은 이야기 속 딸, 며느리, 아내로 설정된 여성인물의 삶에 대한 적극적 공감을 바탕으로 이야기를 전승해온 것이다. 이는 여성화자들의 삶이 ‘가족을 중심’으로 하는 ‘가정 안’의 영역에 한정된 것임을 보여주는 것으로 남성화자들은 여성화자들과 달리 ‘박문수’, ‘중’, ‘정승’ 소재 이야기를 즐겨 구연하고 있다는 사실에서도 드러난다. 즉 남성화자들은 관심의 영역이 집밖으로까지 확대되지만 여성화자들은 가정 안의 관계에 한정된 인식의 차이를 보여준다.

31) 이때 ‘방귀’와 ‘지렁이’를 가족관계를 나타내는 명사로 제시한 것은 이들에 해당하는 이야기가 각각 〈방귀 잘 끼는 며느리〉와 〈지렁이국으로 시어머니 봉양한 며느리〉로 간주해도 무방한 다음의 결과 때문이다. 즉 ‘방귀’를 제목으로 하는 각편 180건의 명사 빈도수를 추출하면 153건이 ‘며느리’로 추출된다. 또한 ‘지렁이’를 제목으로 하는 각편 119건의 명사 빈도수를 추출하면 시어머니(44건), 국(40건), 며느리(38건), 효부(33건), 봉양(33건), 눈(24건) 순으로 나타난다.

5. 결론

본고는 현장에서 이야기를 조사하면서, 그리고 조사된 이야기들을 연구하면서 구비설화 향유층들이 가장 좋아하는 이야기들은 무엇일까 하는 오랜 궁금증의 답을 찾아보기 위해 시도된 것이다. 매회 조사 현장에서 형성된 이야기관은 때면 다양하고도 다른 모습을 보여주기 때문에 남성화자들은 전설을 여성화자들은 민담을 주로 구연한다는 장르적 구연 경향성을 경험적으로 짐작할 수 있을 뿐이었다. 물론 다수의 각편을 보유한 유형에 대해서는 개별 유형의 연구들을 검토함으로써 확인해볼 수는 있다. 하지만 『한국구비문학대계』와 『증편 한국구비문학대계』에 수집된 모든 구비설화 가운데 어떠한 이야기들이 집중적으로 구연되었는지, 남녀화자들은 어떠한 이야기들을 특히 관심 있게 구연하였는지에 대해 객관적 지표를 통해 확인할 수는 없었던 것이다.

본고는 바로 이를 해명하기 위해 ‘디지털 방법론’ 가운데 ‘텍스트마이닝 기법’을 사용하여 이야기의 소재 및 주제를 보여주는 핵심어의 빈도수를 수치화함으로써 많이 수집된 이야기가 무엇인지, 즉 전승집단이 어떠한 이야기를 관심 있게 전승하였는지를 살펴본 것이다. 물론 디지털 방법론과는 별개로 화자들이 집중적으로 구연한 다양한 유형의 이야기들을 제시하다 보니 각 유형을 관심 있게 구연한 이유에 대해서는 구체적인 분석이 이루어지지 못한 것은 본고의 한계이다. 이에 대해서는 후속 논고를 마련하고자 한다.

고전문학 분야에서도 텍스트마이닝 기법을 활용한 연구가 활발하게 진행되고 있는 상황이다. 디지털 기법을 사용한 고전 텍스트 연구가 축적되고 있지만 연구 대상과 논의 방향에 적합한 디지털 분석 모델을 적용하는데 대한 어려움은 고전문학 디지털 연구의 한계로 지적되고 있다. 그럼에도 디지털 기법을 활용한 연구의 필요성은 “컴퓨터를 이용한 연구가 우

리를 미래에 더욱 다가가게 한다”는 에드워드 와그너의 말을 빌리지 않더라도³²⁾ 이전에는 불가능했던 고전문학 연구의 영역을 확장해주는 가능성에 있는 것이다. 즉 디지털 분석 기법을 통해 방대한 고전 텍스트를 거시적 관점에서 분석하고, 다양한 분야의 자료에 접근하여 이를 연구에 활용함으로써 새로운 차원의 연구 방법론으로 고전문학에 대한 해석의 지평을 넓혀 줄 것이라 생각한다.

32) 한국정신문화연구원, 『한국학자료의 전산화연구』, 한국정신문화연구원, 1982; 김현·임영상·김바로, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016, 21쪽 재인용.

참고문헌

1. 자료

<http://gubi.aks.ac.kr>

2. 논저

- 강우규·김바로, 「빅데이터와 고전문학 연구방법론」, 『어문논집』 78, 중앙어문학회, 2019, 7~39쪽.
- 김경혜, 「〈도깨비를 이용해 부자되기〉 설화에 나타난 신화성 연구」, 『인문논총』 43, 경남대학교 인문과학연구소, 2017, 189~214쪽.
- 김연창 외, 「텍스트 마이닝을 이용한 지능적 워드클라우드」, 『한국정보통신학회 2019년 춘계 종합학술대회 논문집』, 한국정보통신학회, 2019, 325~326쪽.
- 김영희, 「마을 지형 및 지명 유래담의 공동체 구성력 탐구」, 『비교민속학』 46, 비교민속학회, 2011, 603~641쪽.
- 김현·임영상·김바로, 『디지털 인문학 입문』, 한국외국어대학교 지식출판원, 2016, 1~516쪽.
- 서대석, 『이야기의 의미와 해석』, 세창출판사, 2011, 1~395쪽.
- 심우장·김영원·황치옥, 「한국설화의 네트워크 지형 연구 시론」, 『구비문학연구』 37, 한국구비문학학회, 2013, 73~105쪽.
- 원혜진·이현영·강승식, 「대규모 텍스트 분석을 위한 한국어 형태소 분석기의 실행 성능 비교」, 『2020년 한국컴퓨터종합학술대회 논문집』, 한국정보과학회, 2020, 401~403쪽.
- 임재해, 「설화의 미디어 기능과 지명전설의 인문지리 정보」, 『한민족어문학』 69, 한민족어문학회, 2015, 41~82쪽.
- 장덕순, 『한국설화문학연구』, 박이정, 1995, 1~512쪽.
- 정재민, 「설화에 나타난 한국인의 행복관」, 『한국인의 삶과 구비문학』, 집문당, 2002, 93~109쪽.
- 조동일, 「『한국구비문학대계』 자료 수집과 설화 분류의 원리」, 『한국구비문학대계 별책부록(I)』, 한국정신문화연구원, 1989, 1~18쪽.
- 한국정신문화연구원, 『한국학자료의 전산화연구』, 한국정신문화연구원, 1982, 1~148쪽.

ABSTRACT

Narrative interest and meaning of oral folktales transmission
group using text mining technique

-For the digital archive of <Korea Oral Literature Daegye> -

Han, Yu-jin

This paper examines the aspect and meaning of narrative interests of the whole tradition group and the tradition group according to gender by using text mining technique targeting each 26,542 oral narratives in the digital archive of <Korea Oral Literature Daegye> . For this purpose, the nouns were extracted by analyzing the morphemes from the titles of 26,542 oral narratives using the Mecab class of the KoNLPy package, and then the high frequencies were extracted using the Counter class of the collection package.

As a result, looking at the 30 nouns occupying the highest frequency, the oral folktale tradition group consists of 'the story of the origin of the place name', the tale of the 'tiger', the story of the 'goblin', the narrative showing the 'relationship of family relationships', 'wealth, filial piety, fortune, it is confirmed that he mainly told stories about the values pursued by humans', such as famous places.

The result of the most oral tales of 'place name origin' shows that the narrators perceive 'story that is worth investigating' as a place name origin story. The story of 'Tiger' brought laughter to the enjoyment class and revealed the value of 'filial piety'. The story of 'Goblin' is a means to satisfy 'story pleasure' by conveying the fictional situation in the story as if it were a real event and arousing interest to the enjoyment class to feel 'creep' or to solve problems that are difficult to explain in reality. it became. Among the stories that show 'family relations', in particular, the story about 'daughter-in-law' was overwhelmingly told, which shows that the family relationship that causes the most conflict in reality is

that surrounding the daughter-in-law. Nouns such as 'son', 'myeong-dang', 'filial piety', 'rich man', 'grave', 'blessed', 'filial piety', etc. It shows that he cares about 'filial piety' above all other ideologies. The character's earnest wish for this is shown to be realized through 'Myeongdang', and it can be interpreted in this context that many stories about 'myongdang' and 'grave' are handed down.

Meanwhile, narrative interest according to gender was examined through 15,088 stories told by 2,916 male narrators and 7,467 stories narrated by 2,004 female narrators. Male narrators recounted many stories about the origin of the place name, and female narrators enjoyed telling stories that showed problems between families. This shows that the lives of female speakers are focused on 'in the home' with a 'family', which is different from male speakers who enjoy telling stories about 'Park Moon-soo', 'Middle', and 'Jeong-seung' and expand their area of interest outside the house. will be.

Key Words

Oral tale, 『Korean Oral Literature Daegye』, 〈Korean Oral Literary Daegye〉 Digital Archive, Digital Methodology, Text Mining, Tradition Group, Story Interest, Origin of Place Name, Tiger, Goblin

논문투고일: 2022.07.13.
 심사완료일: 2022.08.05.
 게재확정일: 2022.08.08.