

## 텍스트 마이닝을 통한 언해본 〈태상감응편〉 감응담의 메시지 분석

김수연\* · 하은하\*\* · 김수민\*\*\* · 김현진\*\*\*\*

### 〈차 례〉

1. 시작하며
2. 〈태상감응편〉 연구사와 질문의 설정
3. 〈태상감응편〉의 구성과 데이터 분석의 과정
  - 1) 〈태상감응편〉의 구성
  - 2) 〈태상감응편〉의 데이터 분석 과정
4. 〈태상감응편〉 감응담의 어휘 성격과 메시지 분석
  - 1) 텍스트 구성 어휘의 성격
  - 2) 텍스트 구성 어휘의 주제 지향
5. 마치며

### 〈국문초록〉

본 연구의 목적은 텍스트 마이닝을 통해 19세기에 조선 왕실에서 간행한 언해본 〈태상감응편〉의 메시지 성격을 분석하는 것이다. 〈태상감응편〉은 국가가 발행한 권선서라는 점에서, 특별한 목적성을 지닌다. 또한 표제와 장르 규정은 ‘도교, 초월자, 선악, 인과’를 중심으로 한 표층 메시지를 구성한다. 선명한 표층 메시지는 독서 과정에서, 인지편향을 작동시킬 우려가 있다. 데이터 마이닝은 인간의 주의력이 놓치기 쉬운 미세 단위 텍스트 정보를 인간의 인지가 파악하는 단위의 정보와 동등한 가치로 다룬다는 점에서, 인지편향이 작동하기 쉬운 텍스트에 대한 새로운 해석을 가능하게 한다.

\* 서울여자대학교 국어국문학과 조교수(주저자)

\*\* 서울여자대학교 국어국문학과 교수(교신저자)

\*\*\* 서울여자대학교 컴퓨터학과 학사(공동저자)

\*\*\*\* 서울여자대학원 국어국문학과 석사수료(공동저자)

이 글에서는 표층 메시지가 강한 언해본 〈태상감응편〉 감응담이 수신자에게 실질적으로 어떠한 메시지를 전달하는지 알아보기 위해 어휘의 빈도분석, 어휘 성격의 시각화, 주제 지향의 시각화를 진행했다. 언어적 접근과 시각화를 위해서는 워드 클라우드, TF-IDF, word2Vec, K-means 클러스터링, 토픽 모델링 등의 정량적 분석기법을 활용했다. 〈태상감응편〉에 사용된 어휘를 중심으로 텍스트 마이닝을 진행한 결과, 〈태상감응편〉이 강조하는 윤리는 종교적 윤리가 아니라 일상의 윤리이고, 감응담이 충족시키는 욕망은 초월적 욕망이 아니라 세속적·일상적 결핐을 채우려는 욕망임을 확인했다. 이것은 기존의 질적 연구가 분석한 결과와 구분된다는 점에서, 새로운 통찰의 가능성을 제공한다.

□ 텍스트 마이닝, 〈태상감응편〉, 메시지 분석, TF-IDF, word2Vec, K-means 클러스터링, 토픽 모델링

## 1. 시작하며

본 연구의 목적은 텍스트 마이닝 솔루션을 통해 19세기 조선 왕실에서 간행한 언해본 〈태상감응편〉의 메시지 성격을 분석하는 데 있다. 텍스트 마이닝은 자연어 처리를 기반으로 하는 데이터 마이닝의 기법 중 하나로서, 텍스트에서 추출한 데이터를 기반으로 의미 있는 통찰을 획득하는 방법이다.

그동안 데이터 분석에 기반한 고전문학 분야의 연구는 양적인 측면에서 적지 않은 성과를 보였다.<sup>1)</sup> 주로 디지털 기반의 방법론을 통해 언어 및

1) 문학분야에서 디지털과 인문학의 융합연구는 다양하게 이루어지고 있다. 여기에서는 고전문학 분야의 연구만 간략히 제시했다. 강우규·김바로, 「컴퓨터를 활용한 〈소현성록〉 연작의 변이양상 고찰-이대15권본과 서울대21권본을 중심으로-」, 『국제어문학회학술대회자료집』 3, 국제어문학회, 2018, 149~159쪽; 강우규·김바로, 「〈소현성록〉 연작에 나타난 감정의 출현 빈도와 의미-컴퓨터를 활용한 통계학적 분석을 바탕으로-」, 『온지논총』 56, 2018, 101~128쪽; 강우규·김바로, 「디지털 맵핑을 통한 〈구운몽〉 연구 및 교육적 활용」, 『열상고전연구』 69, 열상고전연구회, 2019, 217

감정분석, 네트워크분석, 공간분석을 시도했고, 그 과정에서 고전문학에 적용 가능한 디지털 분석 기법을 개발하여, 데이터 모델의 구축을 통한 지식 생산과 활용방안을 마련할 수 있는 토대를 제공했다. 동시에 디지털 방법론을 활용한 분석에 필수적인 토대 데이터 설계와 빅데이터 구축이 미비한 한계도 확인했다.<sup>2)</sup>

빅데이터 구축 문제 다음으로, 고전문학 대상 데이터 분석에서 중요한 문제는 ‘결과의 의미’이다. 기존 연구가 매우 어려운 환경에서 분석을 진행했지만, 많은 연구자들은 그 결과가 갖는 의미에 대해서 회의적이다. 중요한 이유는 분석의 결과가 이전 질적 연구의 결과를 확인하는 데 머물 뿐 새로운 통찰을 끌어내지 못하고 있다는 것이다.<sup>3)</sup>

이와 같은 문제제기는 새로운 방법론의 도입 초기에 나타나는 부득이하면서도 필수적인 현상이다. 질적 연구의 성과를 확인하는 과정에서 양적 방법론의 타당성이 검증되기 때문이다. 그러나 이 단계에 머물러 있어서는 안 되기 때문에, 연구자는 새로운 통찰을 이끌어 낼 방법을 고민해야 한다. 그 시작이 ‘문제의 설정과 질문의 발견’이다. 새로운 방법론이 이전의 방법론과 차별적으로 해결할 수 있는 질문을 던지면서 연구를 시작하는 것이다. 질적 연구에 적합한 질문에 양적 방법론을 적용한다면, 그 결과는 당연히 기존 연구의 확인에 머물 수밖에 없을 것이기 때문이다.

그렇다면 데이터 분석이 의미를 가질 수 있는 질문은 무엇일까? 이에

---

~247쪽; 김성문·김바로, 「디지털 분석 기법을 활용한 시조문학 연구-인공지능(AI) 체제 창작 시조를 중심으로-」, 『문화와융합』 41(5), 한국문화융합학회, 2019, 747~772쪽; 김현, 「디지털 시대의 한문학-데이터로 소통하는 고전 인문 지식-」, 『한문학논집』 49, 근역한문학회, 2018, 9~42쪽.

2) 김바로, 강우규, 「빅데이터와 고전문학 연구방법론」, 『어문논집』 78, 중앙어문학회, 2019, 7~39쪽.

3) 강우규·김바로, 「컴퓨터를 활용한 <소현성록> 연작의 변이양상 고찰-이대15권본과 서울대21권본을 중심으로-」, 『국제어문학회학술대회자료집』 3, 2018, 149~159쪽.

답을 하기 위해서는 데이터 분석의 특징을 생각해볼 필요가 있다. 데이터 분석은 기본적으로 다량의 정보를 다루는 방법론이지만, 동시에 인간의 지각이 소홀히 하는 미세 단위 정보를 기억하는 데 유용한 도구이다. 특히 텍스트 마이닝은 평상시 인간의 주의력이 놓치기 쉬운 미세 단위 텍스트 정보를 인간의 인지가 파악하는 단위의 정보와 동등한 가치로 다룬다. 이에 이 글에서는 텍스트 마이닝 기법을 활용하여 고전 텍스트 중 언해본 〈태상감응편〉의 메시지를 분석해 그동안의 연구에서 포착하지 못한 지점을 발견해 보고자 한다.

## 2. 〈태상감응편〉 연구사와 질문의 설정

〈태상감응편〉은 송나라 때 편찬된 대표적인 도교 권선서이다. 북송조 휘종 정화 8년(1118)에 경문(經文)이 만들어진 후, 크게 3개 계열의 이본이 전승되었다. 첫째는 남송조 이종 재위기(1224-1264)에 편찬된 30권 도장본(道藏本), 둘째는 청조 순치 14년(1657)에 편찬된 8책 도설본(圖說本), 셋째는 청조 강희 56년(1717)에 편찬된 5책 도설본이다. 조선에는 정조 때 이 책이 유입되었고, 철종 3년(1852)에 최성환(1813-1891)이 왕명을 받아 언해본을 초간했으며, 고종 17년(1880)에 중간되었다.<sup>4)</sup>

〈태상감응편〉에 대한 연구는 작품의 개관과 기본 특성, 판본과 전파 과정, 음운학적 특성과 윤리관을 중심으로 이루어졌다. 이능화가 처음 〈태상감응편〉을 소개했고,<sup>5)</sup> 홍윤표가 해제를 썼으며<sup>6)</sup> 오규환, 박성중, 윤찬원,

4) 〈태상감응편〉의 형성과 판본 상황은 김일권, 「장서각 소장 『태상감응편』 언해서의 판본 비교와 권선서의 선복(善福) 인과 윤리사상 고찰」, 『도교문화연구』 55, 2021, 157~220쪽 참조.

5) 이능화, 이종은 역, 『조선도교사』, 보성문화사, 1992, 308쪽.

6) 홍윤표, 「태상감응편도설언해 해제」, 『한국어학자료총서』(전8집) 제3집, 태학사 영

김일권이 관련 연구를 진행했다. 오규환은 <경신록언석>, <태상감응편>, <삼성훈경>, <남궁계적>을 음운학의 관점에서 비교했고<sup>7)</sup> 박성중은 <태상감응편>을 대상으로 도설류 이본을 검토했으며,<sup>8)</sup> 윤찬원은 <태상감응편>의 도교 윤리를 살폈고,<sup>9)</sup> 김일권은 <태상감응편>의 형성과 구성, 편찬과정을 고찰했다.<sup>10)</sup> 대부분은 작품 외적 연구이고, 내용과 주제는 좀더 깊은 천착이 필요한 단계이다.

<태상감응편>의 본질에 다가가는 방법 중 하나는 이 책이 오랫동안 중국은 물론 조선의 민인(民人)에게 수용될 수 있었던 동력을 고찰하는 일이다. 이 점과 관련하여 <태상감응편>이 대부분 ‘이야기’로 구성되어 있다는 점에 주목하고자 한다. 이것은 도교가 ‘이야기 형식’과 매우 긴밀한 종교라는 점과 관련이 있다. 이른 시기부터 공적인 지위를 획득한 유교와 달리, 일견 비합리적으로 보이는 도교는 자신을 존재를 증명하고 신뢰를 확보하는 방안으로 ‘이야기 형식’을 선택했다. 도교는 ‘이성적·합리적 지식’이 아닌 ‘서사적·상상적 지식’에 기반하여 세력을 확장했던 것이다.<sup>11)</sup>

그런데 <태상감응편>의 이야기는 다른 문예물과 달리 국가가 발행한 권선서라는 점에서, 특별한 목적성을 지닌다. 특히 조선의 언해본은 국가가 메시지의 수신자로 일반 백성을 염두에 두었음을 말해준다. 언해리는 형식의 선택은 그 자체로 한글 문해력을 지닌 계층과 소통하려는 행위가 되는

인본, 1986.

7) 오규환, 『『敬信錄 諺釋』과 19세기 후반 도교서의 언해 양상 비교: 『太上感應篇圖說 諺解』·『三聖訓經』·『南宮桂籍』과의 비교를 중심으로』, 『인문학연구』 9(2), 원광대 인문학연구소, 2008, 129~150쪽.

8) 박성중, 『太上感應篇 및 그 圖說類 異本에 대한 管見』, 『민족문화논총』 51, 영남대 민족문화연구소, 2012, 75~104쪽.

9) 윤찬원, 『『太上感應篇』에 나타난 道敎 윤리관 연구』, 『도교문화연구』 31, 2009, 157~183쪽.

10) 김일권(2021), 앞의 논문.

11) 정재서, 『한국도교의 기원과 역사』, 이화여자대학교출판부, 2006, 147~150쪽.

것이다. 정조도 한자를 모르는 지방민에게 유기아와 행걸아의 구휼에 대한 국왕의 진심을 알리고 관심을 부탁하기 위해 언문(諺文)으로 『자훈전칙』을 반포한 바 있다.<sup>12)</sup> 이처럼 국가가 왕의 이름으로 간행한 언해본은 일반 민인을 수신자로 하는 메시지를 담고 있는 것이다.

그렇다면 언해본 〈태상감응편〉의 감응담은 어떠한 메시지를 담고 있는가? 이 질문은 우문(愚問)으로 보인다. “태상감응편”이라는 표제는 이 책이 ‘태상노군이라는 도교의 신격이 인간의 선악 행위를 심판하고 그것에 대한 보응을 통해 인간에게 선을 행하도록 권하는 이야기’라는 분명한 답을 내보이고 있기 때문이다. 또한 ‘권선서’라는 장르 규정은 이 작품의 메시지가 선을 권하는 이념적, 규범적 성격이라는 것을 명시한다. 이처럼 제목과 장르 규정은 ‘도교, 초월자, 선악, 인과’를 중심으로 작품의 표층 메시지를 구성한다.

선명한 표층 메시지는 독서 과정에 적지 않은 영향을 미친다. 일반적으로 인간은 ‘자신에게 최초로 전달되는 정보를 중시’하는 경향이 강하기 때문이다.<sup>13)</sup> 연구자도 이러한 경향에서 자유롭지 못하다면 작품의 분석 과정에서 무의식적으로 표층 메시지에 더 큰 가치를 부여하는 경우도 있을 것이다. 이러한 관점에서 〈태상감응편〉 감응담이 표층 메시지를 그대로 전달하는지, 심층에 또 다른 메시지가 유통되는지를 검토할 필요가 있다.

이 글에서는 언해본 〈태상감응편〉 감응담이 수신자에게 실질적으로 어떠한 메시지를 전달하는지 알아보기 위해 텍스트 마이닝의 방법론을 사용하고자 한다. 구체적으로 감응담에 사용된 어휘를 중심으로 빈도분석, 어휘 성격의 시각화, 주제 지향의 시각화를 진행할 것이다. 언어적 접근과 시각화를 위해서는 워드 클라우드, TF-IDF, word2Vec, K-means 클러스터링, 토픽 모델링 등의 정량적 분석기법을 활용하려 한다. TF-IDF는 텍

12) 김수연, 「조선후기 고소설 속 ‘아동 서사’의 문학적 장면과 사회적 의미-〈창선감의록〉과 〈완월회맹연〉을 중심으로-」, 『한국고전연구』 47, 2019, 30쪽.

13) 홍명희, 『상상력과 가스통 바슬라르』, 살림, 2005, 63쪽.

스트에서 사용된 어휘의 가중치를 도출하는 방법으로, <태상감응편>이 강조하는 키워드가 각 편별로 어떠한 비중을 차지하는지 확인하는 데 유용하다. word2Vec은 텍스트에서 도출된 어휘들을 유의어로 분류하여 <태상감응편>이 중요하게 생각하는 인간관계와 선악보응의 구체적 양상을 계량적으로 보여줄 것이다. K-means 클러스터링은 유사한 성격의 표현들을 군집화하여 전체 메시지를 구성하는 집단의 특성을 확인하는 데에 기여한다. 이러한 분석기법은 감응담이 강조하는 핵심이 무엇인지를 객관적 수치로서 뒷받침할 수 있도록 도울 것이다. 토픽 모델링은 각편 이야기의 메시지 성격을 규명하여, 권선서의 간행 의도가 실질적으로 효과가 있는지를 검토하는 데에 유용한 통찰을 줄 것이다.

연구 텍스트는 장서각 소장 5책본 『감응편(感應篇)』(고종 중간본)을 대상으로 했다. 10책본은 필사본으로 한문 원문과 도설이 없고 내용상의 오류가 적지 않은 데 반해 5책본은 왕명에 의해 여러 차례 정식 간행된 판본으로, 독자층의 확장력이 높은 판본이기 때문이다.

### 3. <태상감응편>의 구성과 데이터 분석의 과정

#### 1) <태상감응편>의 구성

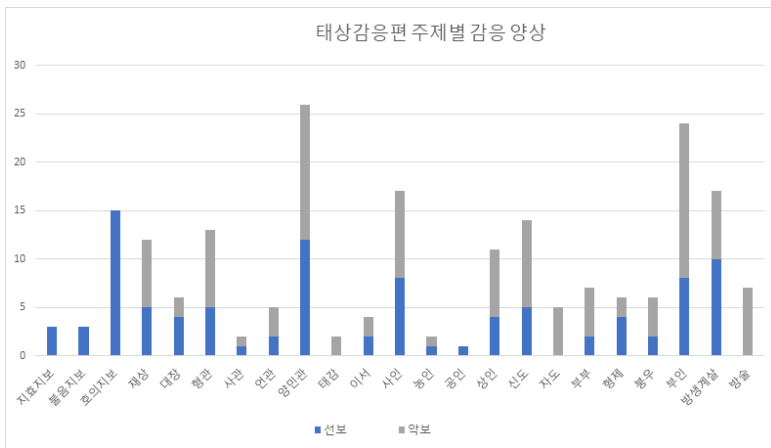
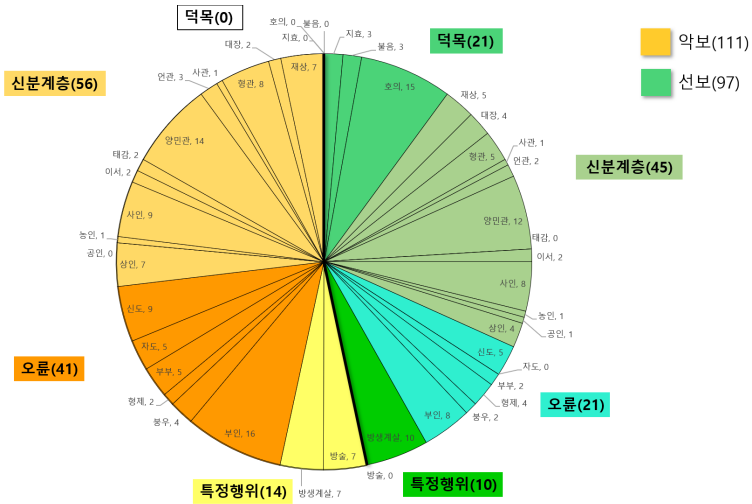
<태상감응편>의 핵심은 감응담이다. 5책본의 감응담은 「감응기(感應記)」에 수록되어 있다. 「감응기」에는 총 206명에 대한 208개의 감응 사례가 있고, 그것은 주제에 따라 39개의 절목으로 나뉜다. 구체적 주제는 1. 지효지보(至孝之報), 2. 불음지보(不淫之報), 3. 호의지보(好義之報), 4. 재상선보(宰相善報), 5. 재상악보(宰相惡報), 6. 대장선보(大將善報), 7. 대장악보(大將惡報), 8. 형관선보(刑官善報), 9. 형관악보(刑官惡報), 10. 사관선보(史官善報), 11. 사관악보(史官惡報), 12. 언관선보(言官善報), 13. 언

관악보(言官惡報), 14. 양민관선보(養民官善報), 15. 양민관악보(養民官惡報), 16. 태감악보(太監惡報), 17. 이서선보(吏書善報), 18. 이서악보(吏書惡報), 19. 사인선보(士人善報), 20. 사인악보(士人惡報), 21. 농인선보(農人善報), 22. 농인악보(農人惡報), 23. 공인선보(工人善報), 24. 상인선보(商人善報), 25. 상인악보(商人惡報), 26. 신도선보(臣道善報), 27. 신도악보(臣道惡報), 28. 자도악보(子道惡報), 29. 부부선보(夫婦善報), 30. 부부악보(夫婦惡報), 31. 형제선보(兄弟善報), 32. 형제악보(兄弟惡報), 33. 봉우선보(朋友善報), 34. 봉우악보(朋友惡報), 35. 부인선보(婦人善報), 36. 부인악보(婦人惡報), 37. 방생계살선보(放生戒殺善報), 38. 방생계살악보(放生戒殺惡報), 39. 방술악보(方術惡報)이다.

1번에서 3번은 제목에 선악이 드러나지 않지만 내용상 선보이다. 지효, 불음, 호의는 각각 지극한 효성의 보응, 음란하지 않은 것에 대한 보응, 의를 좋아한 것에 대한 보응을 나타낸다. 4번부터 39번까지는 선보와 악보를 나누었다. 태감의 악보, 공인의 선보, 자도의 악보, 방술의 악보를 제외하면 대체로 동일 대상의 선보와 악보를 짝지어 제시했다. 그 중 4-25번은 신분 에 따른 선보와 악보를 보였고, 26-36번은 관계에 따른 선보와 악보를 보 였다. 신도는 군신유의(君臣有義), 자도는 부자유친(父子有親)과 관련 있 는 항목이므로 전체가 오륜(五倫)에 의한 관계임을 알 수 있다. 특이한 것 은 부부 외에 별도로 부인의 항목이 추가된 점이다. 나머지는 방생계살과 방술 같은 특정 행위에 대한 보응을 말했다. 작품의 편제를 간략히 나타내 면 다음(69쪽 도표)과 같다.

이와 같은 구성은 <태상감응편>이 덕목, 신분계층, 오륜, 특정 행위라는 4개의 대분류로 구성되어 있고, 전체적으로는 선보보다 악보를 조금 더 강 조하고 있음을 보여준다. 세부적으로는 신분계층, 오륜, 특정 행위에서 악 보가 선보보다 중시되고 있고, 맨 앞에 놓인 덕목은 선보만으로 이루어져 있으며, 특정 행위의 하위 분류이자 편제상 가장 뒤에 놓인 방술은 악보만

으로 구성되어 있다. 이것은 작품의 초기인상을 선보로 구성하고 방술에 대해서는 부정적 이미지를 만들어내는 배치이다.



이것은 도교권선서 <태상감응편>의 감응담 구성 체계가 도교의 초월성·종교성보다 권선의 이미지를 중시하고, 각편의 이야기에서는 선보를 통

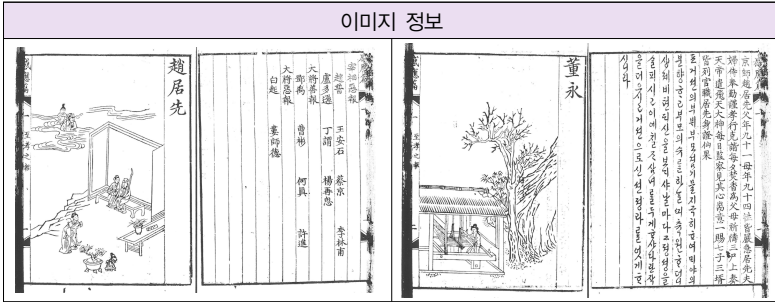
한 권선보다는 악보를 통한 징악을 자극하고 있음을 드러낸다. 선보를 통한 권선은 평화로운 시대에 더 나은 수준의 민심을 고양하기 위해 선을 증진시키는 것이 목적이고, 악보를 통한 징악은 증가하는 악행을 예방하는 것 혹은 통제하는 것이 목적이다. 이러한 점을 고려한다면 <태상감응편>은 기본적으로 혼란한 민심과 사회를 전제로 구성·간행된 것임을 알 수 있다.

## 2) <태상감응편>의 데이터 분석 과정

<태상감응편> 감응답의 데이터 분석 첫단계는 텍스트 분석을 위한 전처리이다. 전처리는 본격적 분석에 앞서, 자료를 컴퓨터가 읽을 수 있는 형태로 정비하는 일련의 활동을 의미한다. 디지털이징, 불용어 처리, 형태소 분석을 포함한 데이터 검증이 여기에 해당한다.

현재 언해본은 pdf 이미지로 제공되고 있다. 이것을 분석 가능한 형태로 변환하기 위해서는 디지털이징(digitizing) 작업을 진행해야 한다. 디지털이징은 아날로그 문서를 디지털로 scan하는 것을 포함하여 디지털 사본이 지속적으로 이용될 수 있게 하는 모든 활동을 가리킨다. 현재 제공되는 pdf 이미지는 종이 원본을 1차 디지털이징한 상태인데, 이것을 다시 컴퓨터 프로그램이 읽을 수 있는 형식으로 변환하기 위해서 2차 디지털이징이 필요한 것이다.

2차 디지털이징에는 한글 고어를 언어 프로그램이 읽을 수 있는 형태로 입력하는 것, 띄어쓰기를 반영하고 표기 형태를 수정하는 것이 포함된다. 현재 수준의 텍스트 마이닝 솔루션으로는 띄어쓰기가 반영되지 않은 한글 고어를 바로 분석할 수 없기 때문이다. 현대 문법을 기준으로 띄어쓰기를 반영하고, ‘.’와 이중모음, 주격 조사 ‘ㅣ’ 등의 고어를 현대 표기로 수정하여 뜻이 통하도록 한 후, 필요한 추가 분석에 대비해 한자를 병기한 데이터를 하나 더 구축했다. 단, 문장구조는 최대한 언해문의 원래 형태를 유지했다. 그 예시는 다음과 같다.



**디지털타이징**

조거선의부부부모섬기물지극히허여미야의분향하고부모의수를하늘찌축원하더니상제비천대신을보니샤날마다그정성을살피시고이에칠자삼녀를두게하시다관작을더하시고거선으로신선정과를얻게하시니라

**띄어쓰기 반영**

조거선의 부부가 부모 섬김을 지극히 하여 매야(每夜)에 분향(焚香)하고 부모의 수(壽)를 하늘찌 축원하더니 상제(天帝)가 비천대신(飛天大神)을 보내사 날마다 그 정성을 살피시고 이에 칠자삼녀(七子三壻)를 두게 하시 다 관작을 더하시고 거선으로 신선 정과를 얻게 하시니라.

**한자 병기**

조거선(趙居先)의 부부가 부모 섬김을 지극히 하여 매야(每夜)에 분향(焚香)하고 부모의 수(壽)를 하늘찌 축원하더니 상제(天帝)가 비천대신(飛天大神)을 보내사 날마다 그 정성을 살피시고 이에 칠자삼녀(七子三壻)를 두게 하시 다 관작을 더하시고 거선으로 신선(神仙) 정과(正果)를 얻게 하시니라.

이렇게 정리된 208편 이야기의 자료 규모는 한자 병기 데이터의 경우 97, 736자(공백 제외), 원고지 605.1장이고 한자를 제외한 자료는 69, 455자, 단어수 26,077개이다. 디지털타이징 처리를 완료한 이후에는 워드 클라우드를 사용하여 전체 이야기의 1차 키워드 분석을 진행했다. 1차 키워드의 분석 결과 ‘하고, 하니, 이다’ 등 불용어가 가장 큰 비중으로 나타났다. 이에 불용어를 분리하고, 2차 키워드 분석을 진행했다. 워드 클라우드를 통한 키워드 분석은 어휘의 사용 빈도를 직관적 형태로 제공함으로써 전체 텍스트에 대한 기본 인상을 제공한다는 점에서 의미가 있다.



다음으로 각 어휘가 실제 텍스트에서 어떠한 비중을 차지하는지를 확인하기 위해 형태소 분석을 진행하고 TF-IDF 중요도로 상위 명사를 선별했다. TF는 전체 텍스트에서 특정 어휘가 등장하는 빈도/차지하는 비중을 나타내며, IDF는 208개의 각 문서에서 해당 어휘가 등장하는 빈도/차지하는 비중이다. 이 두 가지 요소를 중심으로 진행되는 TF-IDF는 전체 문서와 개별 문서 사이의 빈도와 비율을 통해 작품에서 차지하는 해당 단어의 중요도를 표시한다. 분석 자료의 TF-IDF 분석의 결과는 최소 단위로 설정해도 3,300개가 넘는다. 여기에서는 명사와 동사의 상위 10개씩을 제시했다.

명사 상위 10개			동사 상위 10개		
문서번호	index	tf-idf			
1213	사람	12.38649	605	죽다	16.71639
1668	아들	6.59516	221	되다	13.64385
1044	벼슬	5.213849	360	보다	12.06125
1003	백성	4.31306	603	주다	10.1729
2241	일일	3.474518	547	이르다	9.949117
1676	아이	3.463387	606	죽이다	9.676833
692	도적	3.393122	1	가다	9.420103
2666	지아비	3.01571	501	얻다	8.362204
784	마음	2.998462	359	보내다	8.216557
386	그대	2.828089	332	받다	7.916343

워드 클라우드와 TF-IDF 분석을 통해 확인되는 것은 어휘 비중에서 표층 메시지를 구성하는 ‘도교, 초월자, 선악, 인과’의 요소가 미약하다는 점이다. 워드 클라우드의 결과는 직관적으로 이 작품이 도교 권선서라는 정보를 제공하지 않는다. ‘사람이, 땅에, 어찌, 또한, 집에, 일을, 이미’ 등이 가장 눈에 띄고, 도교적 초월성을 짐작할 수 있는 단어 중 빈도가 높은 것은 ‘홀연, 꿈에’이며, 권선과 연결할 수 있는 것은 ‘마땅히’ 정도이다.

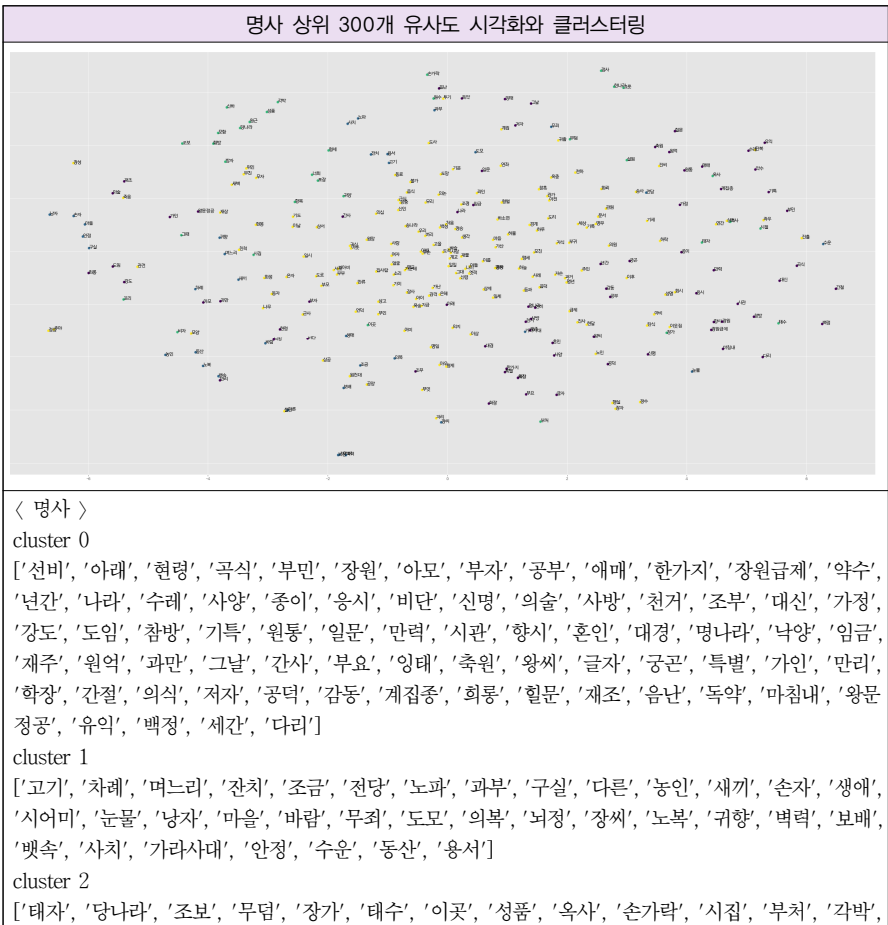
TF-IDF 분석의 결과도 비슷하다. 명사의 경우 상위 10개에 직관적으로 ‘도교, 초월자, 선악, 인과’와 연결할 수 있는 키워드는 보이지 않는다. 동사의 경우는 상대적으로 도교적 권선과 선악보응에 연결할 수 있는 단어가 나타난다. ‘죽다, 죽이다, 얻다, 받다’가 이에 해당한다. 물론 이것도 직관적으로는 도교적 초월관을 드러낸다거나 선악인과를 바로 제시하지는 않는다. 부사의 경우는 ‘홀연, 마땅히’ 정도가 확인되고, 형용사는 ‘있다, 없다, 당하다, 귀하다’ 등을 연결해볼 수 있으나 이 또한 표층 메시지와 바로 연결되지는 않는다.

#### 4. <태상감응편> 감응담의 어휘 성격과 메시지 분석

##### 1) 텍스트 구성 어휘의 성격

워드 클라우드와 TF-IDF를 사용한 1차 어휘 빈도 분석에서는 <태상감응편>에서 일반적으로 기대되는 메시지가 확인되지 않았다. 이에 <태상감응편>의 실질 메시지가 무엇인지를 확인하기 위해 텍스트에서 비중 있게 쓰인 어휘의 기본 성격을 파악했다. 구체적으로 TF-IDF에서 도출된 명사 상위 300개, 동사 상위 300개, 형용사 상위 300개, 부사 상위 300개를 대상으로 word2Vec으로 유사도를 측정하고 후 K-means 클러스터링을 사용하여 군집화를 시행했다.

이때 명사와 동사는 사이킷런(sklearn의 tf-idf이용) min\_df를 2로 설정하여서 단어가 포함된 문서 수를 설정해주었다. 형용사와 부사는 개수가 적어서 min\_df를 설정하지 않았다. word2vec은 skip-gram기법을 사용했고, 앞뒤로 고려하는 단어의 개수 수준은 window=5로 설정했으며, 반복횟수는 iter=500으로 지정했다. 벡터 시각화를 위해서는 t-SNE을 사용해 2차원으로 축소하였다. 여기에서는 명사의 분석을 예로 들었다.



‘정사’, ‘한나라’, ‘비자’, ‘그때’, ‘너희’, ‘혜방’, ‘형제’, ‘항복’, ‘설원’, ‘시절’, ‘포장’, ‘망자’, ‘모함’, ‘귀양’, ‘원수’, ‘그릇’, ‘원근’, ‘신하’, ‘어사’, ‘포리’]  
 cluster 3  
 [‘사람’, ‘아들’, ‘벼슬’, ‘일일’, ‘마음’, ‘백성’, ‘도적’, ‘이름’, ‘고을’, ‘그대’, ‘지아버’, ‘생각’, ‘나이’, ‘아이’, ‘아전’, ‘하늘’, ‘소리’, ‘서로’, ‘신령’, ‘이때’, ‘부인’, ‘우리’, ‘자손’, ‘가난’, ‘장사’, ‘조정’, ‘머리’, ‘귀신’, ‘어미’, ‘송나라’, ‘연고’, ‘상고’, ‘관가’, ‘계집’, ‘아버’, ‘죄인’, ‘가산’, ‘문서’, ‘기록’, ‘옛적’, ‘진씨’, ‘과거’, ‘부모’, ‘주인’, ‘자식’, ‘재물’, ‘무슨’, ‘경계’, ‘경사’, ‘군사’, ‘형벌’, ‘아우’, ‘여자’, ‘평생’, ‘허물’, ‘정승’, ‘무리’, ‘관원’, ‘사랑’, ‘노인’, ‘가운데’, ‘부친’, ‘급제’, ‘집사람’, ‘모양’, ‘부부’, ‘계교’, ‘사례’, ‘명부’, ‘처음’, ‘나무’, ‘등과’, ‘상제’, ‘감격’, ‘의논’, ‘기이’, ‘신인’, ‘진사’, ‘음덕’, ‘원망’, ‘회회’, ‘송사’, ‘무자’, ‘이웃’, ‘근심’, ‘은자’, ‘의심’, ‘기도’, ‘이날’, ‘은혜’, ‘옥중’, ‘목숨’, ‘얼굴’, ‘형제’, ‘지금’, ‘재상’, ‘부귀’, ‘맹세’, ‘세상’, ‘상소’, ‘도사’, ‘성명’, ‘도망’, ‘모친’, ‘도로’, ‘일시’, ‘회시’, ‘연간’, ‘허락’, ‘위인’, ‘하루’, ‘명년’, ‘관전’, ‘친척’, ‘공중’, ‘장수’, ‘도리’, ‘음식’, ‘이후’, ‘현달’, ‘탄식’, ‘투기’, ‘장자’, ‘상공’, ‘언덕’, ‘이웃집’, ‘진홀’, ‘동자’, ‘무엇’, ‘물건’, ‘회생’, ‘기운’, ‘정성’, ‘행실’, ‘명일’, ‘이상’, ‘현몽’, ‘새벽’, ‘동료’, ‘불가’, ‘하소연’, ‘연좌’, ‘좌우’, ‘참혹’, ‘장성’, ‘공양’, ‘기세’, ‘상서’, ‘죽음’, ‘천하’, ‘탄복’, ‘귀줄’, ‘자리’, ‘의원’, ‘원컨대’, ‘의지’, ‘일세’, ‘주야’, ‘만류’]

위의 결과를 대상으로 어휘의 성격을 분석하기 위해 하위 항목을 설정했다. 명사는 각 그룹의 어휘들을 주체 성격의 것과 대상 성격의 것으로 재분류할 수 있다. 재분류 과정에서 텍스트 마이닝 중의 오류를 점검하는 후처리도 병행했다.

명사 제1 그룹의 경우 주체의 성격은 대다수가 선비와 관료이고, 대상의 성격은 시험과 급제 등 사회적 지위를 얻는 관문과 관련이 있다. 대상의 경우, ‘공부, 응시, 천거, 기특, 원통, 향시, 재주, 축원, 공덕, 감동’ 등 다수가 주체에게 주어지는 선보에 대한 원인을 구성하고, ‘간사, 희롱, 음난’ 등 부분적으로 악보에 대한 경계(警戒)를 담았다. 이 중 ‘신명, 축원’ 같은 초월적 성격의 단어가 간간히 보인다. 전체적으로 이 그룹의 어휘들은 사회적 지위를 얻는 과정과 요인을 중심으로 이야기를 구성하는 데 참여하며, 그 과정에서 개인적 성실과 정성을 중시하는 메시지를 전달하는 데에 기여하고 있음을 알 수 있다.

명사 제2 그룹의 경우 주체의 성격은 대다수가 일반 백성이고 그 중 하층 여성이 두드러진다. 대상의 경우, ‘고기, 차례, 잔치, 생애, 눈물, 무죄,

의복, 귀향' 등은 주체의 선보에 대한 원인을 구성하고, '사치' 등이 부분적으로 악보에 대한 경계를 담았다. 이 중 '뇌정, 벽력' 같은 초월성을 추정할 수 있는 단어도 눈에 띈다. 전체적으로 이 그룹의 어휘들은 일반 백성이 생활상의 억울함을 해소하는 과정과 그 요인을 중심으로 이야기를 구성하는 데에 참여하며, 선보 중심 메시지를 전달하는 것에 기여하고 있음을 알 수 있다.

명사 제3 그룹의 경우 주체의 성격은 제1 그룹과 마찬가지로 대다수가 상층의 관료이고, 대상의 성격은 '옥사, 각박, 횡방, 항복, 설원, 모함, 귀양' 등 부귀와 지위를 유지하는 데에 위협을 가하는 시련과 있다. '부처' 같은 초월적 존재도 등장하지만 그 비중은 크지 않다. 전체적으로 이 그룹의 어휘는 세상의 어지러움과 관료 계층의 억울함을 해소하는 과정을 서사화하는 것에 참여하고 악보의 원인을 구성하며, 그 과정에서 선보를 중시하는 메시지를 전달하는 데에 기여하고 있음을 알 수 있다.

명사 제4 그룹은 <태상감응편>에서 가장 큰 비중을 차지한다. 주체의 성격은 대다수가 일반 백성이고 계급적으로 하층에 속한다. 워드 클라우드와 TF-IDF에서 가장 높은 비중을 보인 '사람'을 포함하고 있고, 그 외 단어들은 '아들, 지아비, 아이, 부인, 자손, 어머니, 계집, 아버지, 부모, 자식, 여자, 노인, 집사람, 부부, 이웃, 형제, 모친, 친척, 장자' 등 가족/친족 공동체를 구성하는 어휘가 두드러진다.

대상의 경우는 '벼슬, 가난, 가산, 문서, 과거, 재물, 형벌, 계교, 원망, 회퇴, 송사, 무자, 근심, 의심, 목숨, 부귀, 맹세, 도망, 회시, 진흙, 물건, 희생, 정성, 행실, 하소연, 연좌, 참혹, 공양, 죽음' 등 가족 공동체의 현실적 번영과 실패에 관련된 어휘가 대부분이다. 여기에는 '기이, 음덕, 기도, 현몽' 등 부분적으로 초월적 요소도 보이지만, 비중이 크지 않다. 전체적으로 이 그룹 어휘들은 가족/친족 공동체의 생활과 관련된 번영과 실패의 과정과 그 요인을 중심으로 이야기를 구성하는 데에 참여함으로써 선보와 악보의

원인을 구성하며, 공동체 구성원의 관계적 연결성을 중시하는 메시지를 전달하는 것에 기여하고 있음을 알 수 있다.

동사 그룹의 어휘들은 긍정적 맥락과 부정적 맥락을 지닌 것으로 재분류된다. 동사 제1 그룹의 경우 긍정적 맥락을 구성하는 어휘들은 ‘이끌다, 나누다, 달래다, 권하다, 따르다, 맡다, 갖추다, 감하다, 모이다, 안다, 일으키다, 대신하다, 사하다, 보살피다’ 등 어려움을 함께 하고 고난을 줄여주고, 상호 배려하는 관계 지향의 어휘가 다수이다. 부정적 맥락을 형성하는 어휘로는 ‘늘다, 피하다, 굶다, 다투다, 찌르다, 주리다, 해치다, 빠지다, 업신여기다, 패하다, 태우다, 내치다, 두렵다, 책하다, 부끄럽다, 끼치다’ 등 생활의 곤란을 직접 드러내고, 갈등을 내포하는 단어가 다수를 이룬다.

동사 제2 그룹은 가장 많은 어휘를 포함한다. 특히 긍정적 맥락의 어휘가 집중되어 있다. 긍정적 맥락을 구성하는 어휘들은 ‘되다, 보다, 이르다, 주다, 가다, 얻다, 받다, 알다, 듣다, 행하다, 짓다, 갚다, 구하다, 이루다, 힘쓰다, 웃다’ 등으로 개인적 차원에서 구체적 성취를 이루거나 실천하는 성격이 강하다. 부정적 맥락을 형성하는 어휘로는 ‘죽다, 죽이다, 놀라다, 치다, 빌다, 울다, 꾸짖다, 병들다, 속이다, 앓다, 노하다, 범하다, 갇히다’ 등 생명과 생활의 안정에 위협을 가하는 위기의 속성을 내포하는 단어가 다수를 이룬다.

동사 제3 그룹의 경우, 긍정적 맥락을 구성하는 어휘들은 ‘봉하다, 아끼다, 흥하다’ 등 삶의 조건이 향상되는 이미지를 구성하고, 부정적 맥락을 형성하는 어휘들은 ‘때리다, 파하다, 굶다, 썩다, 당하다, 변하다, 쇠하다’ 등 삶의 조건과 기반이 변화하고 무너지는 이미지를 형성한다. 제4 그룹의 경우, 긍정적 맥락을 구성하는 어휘들은 ‘다스리다, 낫다, 채우다, 바라보다, 모시다.’ 등 공동체의 질서를 유지하거나 회복하는 이미지를 구성하고, 부정적 맥락을 형성하는 어휘들은 ‘빠다, 삭이다, 해하다, 가두다, 빼앗다, 숨다, 상하다’ 등 개인의 신체적·물질적 훼손의 이미지를 만들고 있다.

형용사의 경우는 규범성의 유무와 강약으로 어휘를 재분류할 수 있다. 규범성은 선악(善惡)과 호오(好惡), 시비(是非)와 측은(惻隱)·수오(羞惡)·사양(辭讓) 등과 관련이 있다. 형용사 제1 그룹의 경우, 규범성이 드러나는 어휘들은 강약에 따라 ‘곧다, 부끄럽다, 착한, 미워하다, 좋은, 저어하다’와 ‘촌스럽다, 마지못하다, 유명하다, 황홀한,’ 등으로 나뉘는데, 선행의 구체적 양상을 드러내고, 대상의 행위에 대한 주변의 긍정적 평가와 관련이 있다. 제2 그룹의 경우, 규범성이 드러나는 어휘들은 ‘더럽다, 몹쓸다, 모질다, 흉하다’ 등으로 악행의 구체적인 양상을 드러내고 그 행위에 대한 주변의 응징과 관련이 있다.

형용사 제3 그룹에는 “그르다, 사납다, 분하다, 간사하다, 게으르다, 어지럽다. 부끄럽다, 간사하다“ 등 TF-IDF 분석 결과 형용사 상위 10개의 단어가 모두 포함되어 있다. 이들은 악행의 구체적인 이미지 및 주변의 평가와 관련이 있다. 제4 그룹은 ‘가하다, 기겁다, 아름답다, 귀하다’ 등으로 선행의 구체적인 양상이나, 그에 대한 보응과 관련이 있다.

부사는 운명론적 태도의 유무와 강약을 기준으로 어휘를 재분류할 수 있다. 운명론적 태도는 필연·숙명·천명에 대한 신뢰와 그것에 대한 회의와 의심을 포함한다. 부사 제1 그룹은 ‘감히, 차마, 그제야와 같은 어휘에서 운명에 소극적으로 순응하는 태도가 나타난다. 제2 그룹은, ‘진토록, 종신토록, 기꺼이, 기어이, 해마다, 결단코’와 같은 어휘에서 운명에 대해 적극적으로 순응하는 태도를 드러낸다. 제3 그룹에는 ‘또한, 어찌, 이미, 홀연 일찍, 다시, 마땅히, 이제, 스스로, 모두’ 등 Tf-IDF 분석 결과 부사 상위 단어가 모두 포함되어 있다. ‘어찌, 이미’ 등, 천명의 정당성(응징)을 인정하는 어휘가 다수 나타난다. 제4 그룹은 ‘연즉, 아마, 말미암아, 실로’와 같은 단어로 미루어 볼 때, 거스를 수 없는 운명에 대한 예감(인정)과 관련된 이미지를 만들고 있음을 알 수 있다.



토픽1	토픽2	토픽3	토픽4
사람, 아들, 마음 이름, 생각, 나이 아버지, 죄인, 하늘 과거, 허물, 신령 장원, 아이, 경계 머리, 부친, 평생 자손, 사랑, 서로 공부, 경사, 연고 조보, 우리, 급제	사람, 벼슬, 도적 귀신, 소리, 아들 어미, 백성, 선비 곡식, 아우, 아전 신령, 무덤, 아이 관전, 잔치, 고기 주인, 부인, 얼굴 강도, 자식, 태자 시어미	사람, 백성, 벼슬 문서, 고을, 군사 현령, 장사, 자손 관가, 조정, 관원 아전, 기록, 형벌 나이, 우리, 마음 가산, 집사람, 명부 송사, 생각, 선비 나라, 진사, 노인	사람, 지아비, 계집 부인, 부민, 상고 나무, 은자, 가난 부자, 며느리, 고을 아이, 서로, 감격 연고, 과부, 부모 이름, 회회, 자식 부부, 새끼, 기도 고기, 모양

상위 30개 어휘를 중심으로 4개 토픽의 키워드를 도출하고 각각의 의미를 분석하여 토픽별 주제 지향을 추론했다. 토픽 1에서는 백성 중 부자(父子) 관계 중심의, 조상과 자손 및 가문의 성패와 관련된 주제가 확인된다. 토픽 2는 도적과 강도 등이 등장하는 사회적 혼란상과 그것의 해결과 보응에 대한 주제를 나타낸다. 토픽 3에서는 지배층 관료와 백성 사이의 공정한 법 집행과 관련한 선악 응보(벼슬, 명 잇기)의 주제가 드러난다. 토픽 4는 일반 백성 중 가족 공동체의 여성 구성원을 중심으로, 가정 내 생활상의 고난과 연결된 보응의 주제를 내포한다. 이들을 다시 어휘 성격으로 재분류하면 다음과 같다.

연관 영역	토픽1	토픽2	토픽3	토픽4
일반 백성	사람	사람, 백성	사람, 백성	사람, 상고
지배층	과거, 장원, 조보, 급제	벼슬, 선비, 아전, 관전, 태자	군사, 현령, 관가, 조정, 관원, 아전, 형벌, 송사, 나라, 진사	부자
가족 공동체	아들, 아버지, 부친, 자손	아들, 어머니, 아우, 아이, 부인, 자식, 시어미	자손, 가산, 집사람	지아비, 계집, 부인, 며느리, 아이, 과부, 부모, 자식, 부부
초월적 존재	하늘, 신령	귀신, 신령,	명부	기도

긍정 이미지	사랑, 서로, 경사, 우리	곡식, 잔치, 고기	기록, 생각	감격
부정 이미지	죄인, 허물, 경계	도적, 무덤, 강도	형벌, 송사	회피

재분류 결과는 4개의 토픽이 공통적으로 ‘사람’을 중심으로 한 일반 백성을 주요 키워드로 삼고 있음을 보여준다. 지배층도 백성의 삶과 밀접한 성격을 가진 신분이 강조된다. 가장 많은 서사적 관계는 가족 공동체를 중심으로 구성되며, 긍정적 이미지는 구체적 음식이나 잔치 같은 현실의 일상과 관련이 있고, 부정 이미지는 일상의 질서와 안정에 해를 끼치는 범죄나 송사와 관련이 있다. 초월적 존재는 종교성이 약한 하늘이나 민간 신앙 차원의 신령 등이 확인된다.

## 5. 마치며

“도교적 신들이 태상노군(太上老君)의 명령에 따라서 인간의 행위의 선악 여부를 판정하고 그에 응하여 그 사람의 수명을 줄이거나 늘여주는 강력한 역할을 담당한다는 것이다. 사람은 이와 같은 도덕지침에 따라서 자율적으로 행위하여야 하는 존재이다. 사람이 도덕적으로 살아야 하는 이유는 도덕적 행위를 하는 이유가 옳기 때문만은 아니라 선한 행위를 통하여 도덕적으로 올바른 삶을 구할 수 있을 뿐만 아니라 오래 사는 삶을 얻을 수 있기 때문이다. 도덕적 행위의 끊임없는 축적을 통하여 오래된 삶을 얻는 자가 곧 신선(神仙)이 되는 것이다.”<sup>14)</sup>

인용문은 기존 연구에서 정리한 <태상감응편>의 중심사상이다. 이것은 도교의 인간관이 ‘운명을 선택할 수 있는 자유의지’를 중시한다는 ‘자율적

14) 윤찬원(2009), 앞의 논문, 158쪽.

윤리관'을 강조하기 위한 맥락에서 서술된 것이고, 도교라는 철학적 관점에서 바라본 <태상감응편>의 주제라고 할 수 있다. 이러한 시각에서 추론된 <태상감응편>의 메시지는 선악의 행위에 대해 수명(壽命)의 연장과 단축이라는 응보를 받으며, 그 응보는 결과적으로 장생(長生)을 통해 신선이 되는 과정과 연결됨을 말한다. 즉 '도교'의 불로장생과 신선의 개념을 권선의 맥락으로 구현한 것이 <태상감응편>의 핵심이라고 이해한 것이다.

이러한 결론은 작품의 표층 메시지와 매우 유사하다. 이 경우 작품을 해석하는 과정 중에 인지편향이 작동했을 것이라고 단언할 수는 없지만, 연구자가 중시하는 정보가 추론의 과정에 개입할 가능성도 배제할 수 없다. 신체적 독서에 기반한 분석의 경우, 어떠한 정보에 주목하느냐가 주제적 메시지 도출에 영향을 주기 때문이다. 또한 '이야기'로 전달되는 정보는 철학적 관점보다는 감성적·효용적 맥락으로 수용된다. 감성적·효용적 맥락은 개인마다 다르다. 그러한 까닭에 이야기에 포함된 정보의 가치를 최대한 동등한 위치에 두고, 독자가 이 책을 통해 자극받는 인간의 자율적 선행과 악행의 구체적 내용, 그것의 구성 요소와 결과가 무엇인지를 확인하는 것은 의미가 있다.

본 연구는 이를 위해 텍스트 마이닝을 진행했고, 그 결과 <태상감응편>가 일반 백성의 일상과 밀접한 메시지를 전달하고 있다는 것을 알 수 있었다. 감응담을 구성하는 어휘들은 지배층의 서사라도 백성의 억울함이나 생활상의 안정과 관련된 선악과 연결된다. 이것들은 이 작품이 강조하는 윤리가 종교적 윤리가 아니라 일상의 윤리이고, 감응담이 충족시키는 욕망이 선화(仙化)와 같은 초월적 욕망이 아니라 세속적·일상적 결핍을 채우려는 욕망을 반영한다. 이것은 <태상감응편>의 메시지가 보여주는 장생(長生)이 기존연구가 말한 개인의 신체적 수명 연장이 아니라, 현생에서 안정적으로 생활 기반을 유지하고, 후대 자손에 의해 가문이 이어지는 것임을 나타낸다. <태상감응편> 감응담의 신격(神格)도 '태상'이나 '태상노군' 같

은 종교적 신격보다는 ‘하늘, 신령’ 같은 일상적 수준의 신격이 강조된다. 이들은 철학적·윤리적 맥락보다는 실질적·즉각적 보상과 보응을 위해 논리적 시간을 단축하거나 절차를 생략할 때 소환된다.

텍스트 마이닝을 통해 도출한 <태상감응편>의 메시지는 <삼강행실도>와 같은 유교 권선서와의 비교에서 그 특성이 더 부각된다. <삼강행실도>는 어휘가 관념적이고 추상적이며 보상의 내용이 행위자의 명예를 높이는 방식이라는 점에서, 시선이 위에서 아래로 향하고 있다. 즉 선악보응을 판단하는 최상층의 욕망이나 필요가 반영된 이야기인 것이다. 반면 <태상감응편>은 메시지가 이념적 교화가 아니라 일상을 돌보는 이야기라는 점에서 이 텍스트를 수용하는 백성들의 삶과 밀착되어 있다. 이것이 민간에서 <태상감응편>을 적극 수용한 이유가 될 것이다.<sup>15)</sup>

마지막으로 본 연구는 <태상감응편>의 표층 메시지와 실질 메시지의 거리를 확인했다는 점 외에, 연구자에게 자기 확신에 대한 겸허한 성찰의 기회를 제공했다는 데 의미가 있다. 일반적으로 디지털 연구환경에서 인문학자에게 필요한 것이 분석 기술의 습득이라고 생각하기 쉽다. 그러나 이번 과정에서 가장 어렵고 중요했던 것은 ‘단계마다 어떠한 질문을 던지고 분석 기준을 세우며, 기술적 분석의 결과를 어떻게 해석할 것인가’였다. 글에서는 간단히 한두 줄로 정리되는 분류 기준과 해석이 나오기까지, 수많은 고민과 토론이 필요하다. 이때 가장 중요한 것은 참여 연구자들이 갖추고 있는 질적 연구의 수준과 깊이이다. 의미 있는 양적 연구를 위해서는 작품에 대한 이해를 훈련하는 질적 연구를 심화해야 할 것이다. 디지털 인문학이라는 영역에서 인문학자가 실질적으로 만나는 것은 기술 부족에 대한 부끄러움보다는 자신이 자부했던 인문학적 통찰의 부족에 대한 부끄러움이기 때문이다.

15) <태상감응편>과 <삼강행실도>의 구체적 메시지 분석은 후속 작업으로 보고할 예정이다.

참고문헌

〈태상감응편〉, 장서각 소장본.

- 강우규·김바로, 「〈소현성록〉 연작에 나타난 감정의 출현 빈도와 의미-컴퓨터를 활용한 통계학적 분석을 바탕으로-」, 『온지논총』 56, 2018, 101~128쪽.
- 강우규·김바로, 「컴퓨터를 활용한 〈소현성록〉 연작의 변이양상 고찰-이대15권본과 서울대21권본을 중심으로-」, 『국제어문학회학술대회자료집』 3, 2018, 149~159쪽.
- 김바로, 강우규, 「빅데이터와 고전문학 연구방법론」, 『어문논집』 78, 중앙어문학회, 2019, 7~39쪽.
- 김수연, 「조선후기 고소설 속 '아동 서사'의 문학적 장면과 사회적 의미-〈창선감의록〉과 〈원월회맹연〉을 중심으로-」, 『한국고전연구』 47, 2019, 5~38쪽.
- 김일권, 「장서각 소장 『태상감응편』 언해서의 판본 비교와 권선서의 선복(善福) 인과 윤리사상 고찰」, 『도교문화연구』 55, 2021, 157~220쪽.
- 박성중, 「太上感應篇 및 그 圖說類 異本에 대한 管見」, 『민족문화논총』 51, 영남대 민족문화연구소, 2012, 75~104쪽.
- 오규환, 「『敬信錄諺釋』과 19세기 후반 도교서의 언해 양상 비교: 『太上感應篇圖說諺解』·『三聖訓經』·『南宮桂籍』과의 비교를 중심으로」, 『인문학연구』 9(2), 원광대 인문학연구소, 2008, 129~150쪽.
- 윤찬원, 「『太上感應篇』에 나타난 道教 윤리관 연구」, 『도교문화연구』 31, 2009, 157~183쪽.
- 이능화, 이종은 역, 『조선도교사』, 보성문화사, 1992, 3~511쪽.
- 장 프랑수아 리오타르, 이현복 역, 『포스트모던적 조건』, 서광사, 1992, 5~143쪽.
- 정재서, 『한국도교의 기원과 역사』, 이화여자대학교출판부, 2006, 5~302쪽.
- 홍명희, 『상상력과 가스통 비슬라르』, 살림, 2005, 3~94쪽.
- 홍윤표, 「태상감응편도설언해 해제」, 『한국어학자료총서』(전8집) 제3집, 태학사 영인본, 1986, 1~6쪽.

## ABSTRACT

Analysis of the message of Gam Eung-Story in  
*Taesanggameung-pyeon* through text mining

Kim, Soo-youn &amp; Ha, Eun-ha &amp; Kim, Su-min &amp; Kim, Hyun-jin

This study aims to analyze the message of *Taesanggameung-pyeon* published by the Joseon royal family in the 19th century through text mining. As it was issued by the state, the text of *Taesanggameung-pyeon* occupies a special position. In addition, the title and genre regulation of the book constitute a superficial message centered on "Taoism, the transcendental, good and evil, and cause and effect." Superficial messages may trigger cognitive bias in the reading process. Data mining enables a new interpretation of texts that are prone to cognitive biases in that it treats fine-scale text information, which is easy to miss by human attention, as having the same value as information in units that human cognition readily understands.

In this article, frequency analysis of vocabulary, visualization of vocabulary characteristics, and topic-oriented visualization were conducted to determine the actual message delivered to the readers of *Taesanggameung-pyeon*, which has a strong superficial message. For the linguistic approach and visualization, quantitative analysis techniques such as word cloud, TF-IDF, word2Vec, K-means clustering, and topic modeling were used. Following text mining focused on the vocabulary used in *Taesanggameung-pyeon*, the ethics emphasized in *Taesanggameung-pyeon* is not religion-specific, but has general applicability, and the desire that the Gam Eung Story satisfies is a transcendent desire. Rather, it was confirmed that it was a desire to make up for worldly and everyday shortcomings. This affords new insights in that it is distinct from the results of earlier qualitative studies.

**Key Words**

text mining, *Taesangameung-pyeon*, message analysis, TF-IDF, word2Vec, K-means clustering, topic modeling

논문투고일 : 2022.03.31.

심사완료일 : 2022.05.03.

게재확정일 : 2022.05.03.