

컴퓨터 문헌 분석 기법을 활용한 허균(許筠) 5전(傳)의 서사 단락 구조 비교 연구

김 해 관¹ · 정 화 영² · 김 윤 호³ · 안 영 훈^{4*}

¹경희대학교 국어국문학과 박사과정

²경희대학교 후마니타스칼리지 교수

³목원대학교 컴퓨터융합학부 교수

⁴경희대학교 국어국문학과 교수

A Comparative Study on the Narrative Paragraph Structure of Heo Gyun's Five Biographical Essays Using Computer Literature Analysis

Hae-Kwan Kim¹ · Hwa-Young Jeong² · Yoon-Ho Kim³ · Young-Hoon An^{4*}

¹Ph.D course, Department of Korean Language and Literature, Kyung Hee University, Seoul 02447, Korea

²Professor, Humanitas College, Kyung Hee University, Seoul 02447, Korea

³Professor, School of Computer Convergence, Mokwon University, Daejeon 35349, Korea

⁴Professor, Department of Korean Language and Literature, Kyung Hee University, Seoul 02447, Korea

[요 약]

본 연구는 허균의 『성소부부고』 소재 5전의 서사 단락 구조를 비교하였다. 5편에 공통으로 나타나는 서사 전환점을 기준으로 22개 단락을 부호화하고 해밍 거리 행렬을 산출한 후, 다차원 척도법과 계층적 군집 분석을 적용하였다. 그 결과 엄처사전·손곡산인전·남궁선생전의 유학·현실형과 장산인전·장생전의 신이·탈속형이 구분되었으며, 「남궁선생전」은 두 유형이 교차하는 전환형이었다. 이는 허균이 동일한 전 장르 안에서 유학적·사회적 인물과 도교·신이적 인물을 구별했음을 보인다. 또한 해밍 거리 분석은 이본 뿐 아니라 한 문집의 여러 작품을 비교하고 서사 형식의 공통성을 수치화하는 데에도 활용될 수 있다.

[Abstract]

This study aimed to compare the narrative structures of five biographies in Seongsobubugo by Heo Gyun. To conduct the comparison, 22 narrative segments were coded on the basis of shared narrative turning points, and a Hamming distance matrix was analyzed using multidimensional scaling (MDS) and hierarchical cluster analysis. The results identified two distinct groups: the Confucian–realist type (Eom Cheosajeon, Songgoksaninjeon, and Namgung Seonsaengjeon) and the supernatural–transcendent type (Jangsaninjeon and Jangsaengjeon), with Namgung Seonsaengjeon serving as a transitional case between them. These findings suggest that Heo Gyun differentiated Confucian and socially grounded figures from Daoist and supernatural figures within the same biographical genre. The results also demonstrate that Hamming distance analysis can be applied not only to textual variants but also to comparing multiple works within a single collection and quantifying similarities in narrative form.

색인어 : 허균 5전, 성소부부고, 전(傳), 컴퓨터 문헌 분석 기법, 해밍 거리

Keyword : Heo Gyun 5-Jeon, Seongsobubugo, Jeon, Computer Literature Analysis Method, Hamming Distance

<http://dx.doi.org/10.9728/dcs.2026.27.8.2381>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 07 July 2026; **Revised** 27 July 2026

Accepted 31 July 2026

***Corresponding Author; Young-Hoon An**

Tel:

E-mail: yhnahn@khu.ac.kr

I. 서 론

허균(許筠, 1569~1618)은 조선 후기의 대표적인 문인으로, 그가 편찬한 성소부부고(惺所覆瓿藁)는 시부(詩賦)·문(文)·설(說) 등 방대한 분량을 자편(自編) 체계로 수록한 문집이다. 이 중 문부(文部)에 실린 엄처사전·손곡산인전·장산인전·남궁선생전·장생전의 5편(傳)은 허균이 직접 목도하거나 전언(傳言)으로 접한 이인(異人)·일민(逸民)의 행적을 기록한 전기 산문이다. 이 5편에 관한 선행 연구는 개별 작품의 사상적 분석[1]~[4], 장르론적 접근[5],[6], 허균의 이단 수용 문제[7],[8] 등 다양한 맥락에서 이루어졌으나, 5편을 동일한 분석 틀로 묶어 구조적으로 비교한 연구는 드물다.

최근 고전 문학 연구에 컴퓨터 문헌 분석 기법을 도입하는 시도가 이루어지고 있다[9]~[13]. 특히 김동건·정화영은 해밍 거리(Hamming Distance)를 이용하여 토끼전 이본(異本) 62종의 서사 단락 유사도를 수치화하고 계층적 군집 분석으로 계통을 분류하였다[14]~[16]. 이 방법론은 이본이 다수 존재하는 판소리계 소설에 주로 적용되어 왔으나, 단일 텍스트 내 복수 작품의 서사 구조 비교에도 동일한 논리로 적용할 수 있다.

본 연구는 위 방법론의 핵심 공정을 변용하여, 5편의 전(傳)을 분석 단위로 삼고, 22개의 서사 단락(SID)으로 구조화하여 내용 유형을 인코딩한 뒤 해밍 거리 행렬을 산출하는 시론이다. 분석 결과는 다차원 척도법(MDS)과 계층적 군집 분석을 통해 시각화하며, 이를 바탕으로 허균 5편의 서사 구조적 유형 분화를 규명하는 것이 목적이다. 이는 기존의 텍스트 마이닝 기법으로 다양하게 시도해오던 컴퓨터 문헌 분석 기법과 차별점을 두며, 본 연구에서는 해밍 거리 행렬 산출 및 다차원 척도법을 이용한 시도를 통해 입체적인 해석을 하는 데에 중점을 둔다[17].

허균의 5편 전에 관한 연구는 크게 세 흐름으로 정리된다. 첫째, 사상적 접근이다. 최지선은 남궁선생전을 중심으로 허균의 도교적 지향과 현실 정치 비판을 분석하였으며[1], 양언석은 허균의 불교관 수용을 고찰하였다[7]. 둘째, 문헌론적 접근이다. 이규운은 허균의 문장론과 자편(自編) 체계를 논하였고[5], 임미정은 이본 연구를 시도하였다[8]. 셋째, 디지털 인문학적 접근은 본 연구의 직접 선행 작업에 해당한다. 김동건·정화영의 해밍 거리 기반 이본 계통 분류 방법론이 그것이다[14]. 그러나 허균 5편을 동일한 수치 분석 틀로 비교한 연구는 현재까지 확인되지 않는다.

II. 대상 자료와 서사 단락 분석

2-1 대상 자료

본 연구의 분석 대상은 한국고전종합DB에서 제공하는 성소부부고 소재 5편의 전이다. 이본이 단일한 관계로 이본 계

표 1. 분석 대상 5편의 전(傳) 기본 정보

Table 1. Basic information on the five biographies analyzed

Title of Work	Main Character	Major Status/Background	Relationship with Heo Gyun
Biography of Eom Cheosa	Eom Chung-jeong	A scholar from Gangneung, known for his filial piety	No direct interaction (mentioned in *Oesa-ssi Nonchan)
Biography of Songok Sanin	Lee Dal	A poet of illegitimate birth, from Songok, Wonju	Heo Gyun's poetry teacher; mentioned directly
Biography of Jang Sanin	Jang Han-ung	From a three-generation Yang-ui family; Taoist practitioner	No direct interaction (hearsay)
Biography of Master Namgung	Namgung Du	Yangban from Impi; practitioner after fleeing his home	Met in person in Buan in 1608
Biography of Jang Saeng	Jang Saeng	Son of the Left Magistrate of Miryang; wandering beggar	Close personal friendship in his youth

통 비교가 아닌, 5편 간의 서사 구조 비교를 수행한다.

2-2 서사 단락 분절 및 내용 유형 인코딩

본 연구에서는 5편을 아우르는 공통 서사 구조를 도출하여 22개의 서사 단락(SID)을 설계하였다. 서사 내용 전환점(인물 등장·퇴장, 사건 발생, 장면·공간 이동)을 단락 경계로 삼았으며, 인코딩 기호 체계는 토끼전 연구[14]의 방식을 변용하여 적용하였다. a·b·c·d는 독립적인 내용 유형을, a1·a2는 동일 유형 내 세부 변이를, X는 해당 단락의 부재를 표시한다. 5편에 공통으로 나타나는 단락은 은거·방랑 생활(S14)과 타 인물과의 교류(S13)이며, 이는 5편 모두 이인(異人)·일민(逸民)의 탈속적 삶이라는 장르적 전제를 공유함을 보여 준다.

III. 해밍 거리 기반 서사 구조 유사도 분석

본 연구의 분석 절차는 그림 2와 같이 자료 구조화, 정량 분석, 결과 해석의 세 단계로 구성된다. 먼저 분석 대상인 허균의 5편 전을 선정한 뒤, 작품에 공통으로 적용할 수 있는 서사 전환점을 기준으로 각 작품을 22개의 서사 단락(SID)으로 분절하고, 단락별 내용 유형을 기호화하였다. 다음으로 작품 간 코딩값의 차이를 바탕으로 해밍 거리 행렬을 산출하고, 다차원 척도법(MDS)과 Ward 연결법에 의한 계층적 군집 분석을 통해 작품 간 구조적 유사성과 군집 형성 양상을 시각화하

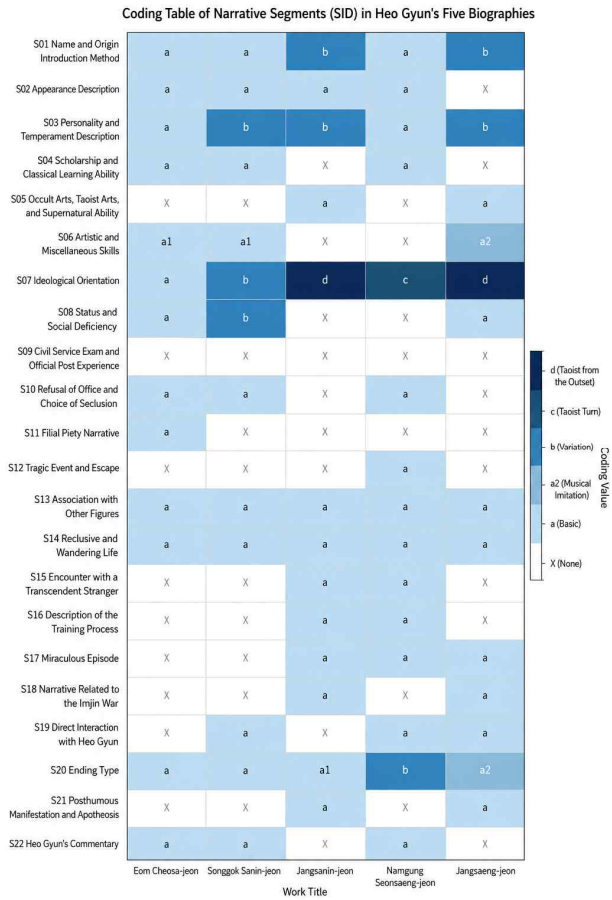


그림 1. 허균 5전 서사 단락 유형 코딩 테이블(SID x 작품)
Fig. 1. Coding table for narrative paragraph types in Heo Gyun's "five essays" (SID x work)

였다. 마지막으로 정량 분석 결과를 토대로 5편을 A·B 두 군집으로 분류하고, 각 군집의 서사적 특성과 함께 두 유형의 성격이 교차하는 전환형 서사 구조를 해석하였다. 이러한 절차는 개별 작품에 대한 질적 독해를 서사 단락 단위의 구조화된 데이터로 전환하고, 정량 분석 결과를 다시 문학적 해석으로 연결하기 위한 분석 틀이다.

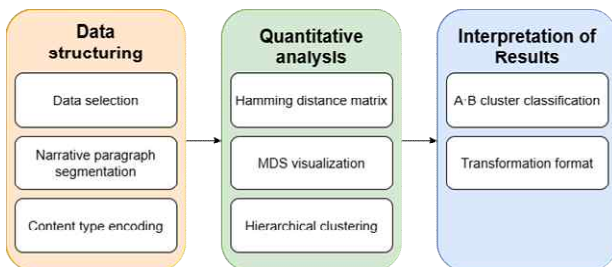


그림 2. 자료수집 과정을 포함한 연구 흐름도
Fig. 2. Flowchart for research including data structuring

표 2. 서사 단락 목록(SID) 및 코딩 기준

Table 2. List of narrative paragraphs (SID) and coding criteria

SID	Narrative Section Unit	Explanation of Encoding Criteria
S01	Introduction of the Protagonist's Name and Origin	a = Explicit, b = Origin Unknown
S02	Physical Description of the Protagonist	a = Present, X = Absent
S03	Description of the Protagonist's Character and Temperament	a = Confucian Character, b = Unconventional Temperament, X = Absent
S04	Scholarly and Confucian Learning	a = Present, X = Absent
S05	Possession of Magical, Taoist, and Supernatural Abilities	a = Present, X = Absent
S06	Artistic and Miscellaneous Talents	a1 = Poetry/Calligraphy, a2 = Music/Imitation, X = Absent
S07	Ideological Orientation	a = Confucianism, b = Poetics/Literature, c = Transition to Taoism, d = Taoism (Innate), X = Absent
S08	Social and Status-Related Disadvantages	a = Poverty, b = Illegitimate or Low Status, c = Abandonment, X = Absent
S09	Success in the Civil Service Examination and Governmental Experience	a = Present, X = Absent
S10	Rejection of Official Office and Reclusion	a = Present, X = Absent
S11	Narrative of Filial Piety	a = Present, X = Absent
S12	Narrative of Tragic Events and Flight	a = Present, X = Absent
S13	Interactions with Other Figures	a = Present (Common to All Five Works)
S14	Life of Reclusion and Wandering	a = Present (Common to All Five Works)
S15	Encounter with a Transcendent Figure	a = Present, X = Absent
S16	Detailed Description of the Cultivation Process	a = Present, X = Absent
S17	Episodes of Supernatural Power Manifestation	a = Present, X = Absent
S18	Narrative Related to the Imjin War	a = Present, X = Absent
S19	Direct Interaction with Heo Gyun	a = Present, X = Absent
S20	Ending Type	a = Natural Death, a1 = Supernatural Death, a2 = Apotheosis, b = Unspecified Ending
S21	Posthumous Manifestation · Apotheosis	a = Present, X = Absent
S22	Heo Gyun's Commentary	a = Present, X = Absent

표 3. 허균 5전 서사 단락 해밍 거리 행렬(평균값)

Table 3. Mean Hamming distance matrix of Heo Gyun's five biographies

	Eom Cheosa-jeon	Songgok Sanin-jeon	Jang Sanin-jeon	Namgung Seonsaeng-jeon	Jangsaeng-Jeon
Eom Cheosa-jeon	0	0.636	2.455	1.636	2.136
Songgok Sanin-jeon	0.636	0	2.364	1.364	1.773
Jang Sanin-jeon	2.455	2.364	0	1.818	1.136
Namgung Seonsaeng-jeon	1.636	1.364	1.818	0	2.545
Jangsaeng-Jeon	2.136	1.773	1.136	2.545	0

3-1 거리 행렬 구성

서사 단락 유형 코딩 테이블을 바탕으로 해밍 거리를 산출하였다. 각 서사 단락의 내용 비교는 다음 척도를 적용하였다: 내용 유형이 동일하면 거리 0, 동일 유형 내 세부 변이(a1 vs a2)는 1, 상이한 유형(a vs b, b vs c 등)은 2, 단락 유무 차이(코드 vs X)는 4이다. 22개 서사 단락의 값을 합산한 뒤 단락 수로 나누어 평균 거리를 산출하면 표 3과 같은 5x5 대칭 행렬이 도출된다.

3-2 다차원 척도법 및 계층적 군집 분석

표 3의 거리 행렬을 다차원 척도법(MDS)으로 2차원 공간에 투사하고, 계층적 군집 분석(Ward 연결법)으로 군집 형성 과정을 시각화한 결과가 그림 2이다.

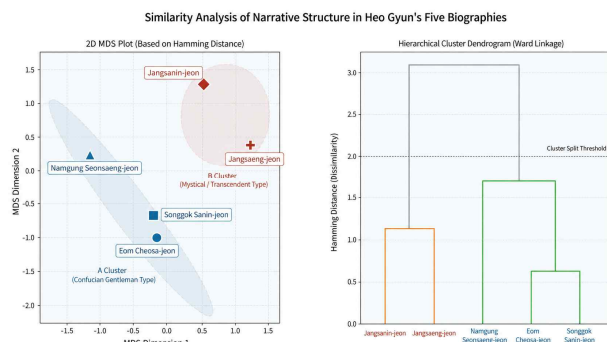


그림 3. 다차원 척도법(MDS) 및 계층적 군집 분석 결과

Fig. 3. Results of multidimensional scaling (MDS) and hierarchical cluster analysis

그림 3의 덴드로그램에서 임계 거리 2.0을 경계로 두 군집이 명확히 분리된다. A군집(엄처사전·손곡산인전·남궁선생전)은 먼저 엄처사전·손곡산인전이 거리 0.636으로 결합하고, 이후 남궁선생전이 합류하는 순서로 형성된다. B군집(장산인전·장생전)은 상호 거리 1.136으로 결합한다. 두 군집 간 거리는 3.090으로, 군집 내부 거리보다 현저히 크다.

Ⅳ. 분석 결과 해석: 두 유형의 서사 전략과 전환형 서사 구조

4-1 A군집: 처사형 서사 문법과 유학적 이인상

엄처사전·손곡산인전·남궁선생전 상호 간의 해밍 거리 최솟값은 0.636(엄처사전-손곡산인전)이다. 이 수치는 5편 전체의 거리 행렬 중 가장 작은 값으로, 두 작품이 서사 구조적으로 매우 유사함을 보여 준다. 두 편은 주인공의 유학적 출신 배경(S01=a), 학문·경학 능력(S04=a), 벼슬 거부(S10=a), 허균의 논찬(S22=a)을 공유하며, 이른바 처사형(處士型) 서사

표 4. 두 군집의 서사 단락 구성 대비

Table 4. A comparison of narrative paragraph structures between the two clusters

Classification Categories	Cluster A (Study Abroad- Realistic Type)	Cluster B (Divine-Transcendent Type)
Target Work	Eom Cheosa Jeon · Songgok Sanin Jeon · Namgung Seonsaeng Jeon	Jang Sanin Jeon · Jang Saeng Jeon
The Protagonist's Background	Clear Introduction to Background and Family History	History unknown or uncertain
Ideological Orientation	Confucianism (a), Poetics (b), Transition to Taoism (c)	Taoism and the Supernatural (d), None (X)
Academic and Governmental Experience	Present in All Three Works	Absent from Both Works
Magic and Divine Abilities	None (X) or Training Results	Innate
Supernatural Episodes	Present Only in Namgung Seonsaeng Jeon	Present in Both Works
Related to the Imjin War	Absent	Present in Both Works
Heo Gyun's Commentary	Present in All Three Works	Absent from Both Works
Ending Type	Natural Death (a) or Unspecified Ending (b)	Supernatural Death (a1) · Apotheosis (a2)

문법을 구성한다. 처사형 서사 문법이란, 제도권 학문을 갖추었음에도 자발적으로 관직을 거부하고 재야에 머문 인물의 행적을 기록하는 일정한 서술 관습을 의미한다. 허균의 논란(S22)이 두 편 모두에 존재한다는 점은, 이 인물들이 허균 자신의 이상적 인간형과 밀접하게 연결되어 있음을 시사한다. 실제로 허균은 엄처사전 말미의 논란에서 처사(處士)는 가정에서 효도를 다했고, 고을에서 절도있는 행실을 하였으니 분

명히 높은 지위에 오를 수 있었다고 평함으로써, 처사형 인물이 자질을 갖추고도 스스로 출사를 자제했음을 직접 확인해 준다.

세 편 사이에서도 미세한 차이가 나타난다. 엄처사전은 효행 서사(S11=a)라는 단락이 유일하게 존재하는 반면, 비극적 사건·도주(S12)와 초월적 이인 만남(S15)이 없어 가장 전통적인 유학 군자형 서사에 가깝다. 이와 달리 남궁선생전은 A

표 5. 남궁선생전 SID별 군집 귀속 분석(분석 제외 S09·S13·S14 제외, 19개 SID)

Table 5. SID-level cluster affiliation analysis of Namgung Seonsaeng-jeon (excluding S09, S13, and S14; 19 SIDs)

SID	Narrative Segment	Code for Namgung Seonsaeng-jeon	Mean Distance to Cluster A	Mean Distance to Cluster B	Cluster Affiliation
S01	Introduction of the Protagonist's Name and Origin	a	0.0	2.0	Cluster A
S02	Physical Description	a	0.0	2.0	Cluster A
S03	Description of Character and Temperament	a	1.0	2.0	Cluster A
S04	Scholarship and Classical Learning	a	0.0	4.0	Cluster A
S05	Occult, Daoist, and Supernatural Abilities	X	0.0	4.0	Cluster A
S10	Refusal of Office and Choice of Reclusion	a	0.0	4.0	Cluster A
S18	Imjin War-Related Narrative	X	0.0	4.0	Cluster A
S21	Posthumous Manifestation · Apotheosis	X	0.0	4.0	Cluster A
S22	Heo Gyun's Commentary	a	0.0	4.0	Cluster A
S06	Artistic and Miscellaneous Talents	X	4.0	2.0	Cluster B
S08	Social and Status-Related Disadvantages	X	4.0	2.0	Cluster B
S11	Narrative of Filial Piety	X	2.0	0.0	Cluster B
S15	Encounter with a Transcendent Figure	a	4.0	2.0	Cluster B
S16	Cultivation Process	a	4.0	2.0	Cluster B
S17	Episodes of Supernatural Manifestation	a	4.0	0.0	Cluster B
S07	Ideological Orientation	c (Transition to Taoism)	2.0	2.0	Neutral (Unique)
S12	Tragic Events and Flight	a	4.0	4.0	Neutral (Unique)
S19	Direct Interaction with Heo Gyun	a	2.0	2.0	Neutral
S20	Ending Type	b (Unspecified Ending)	2.0	2.0	Neutral (Unique)

Note. The distance to Cluster A is the mean distance between the code assigned to Namgung Seonsaeng-jeon and the corresponding codes assigned to Eom Cheosa-jeon and Songgok Sanin-jeon. The distance to Cluster B is the mean distance between the code assigned to Namgung Seonsaeng-jeon and the corresponding codes assigned to Jang Sanin-jeon and Jangsaeng-jeon. Distances were assigned as follows: identical codes = 0; within-type variations (a1 vs. a2) = 1; different content types = 2; and presence-absence differences (a code vs. X) = 4. "Neutral" indicates an equal mean distance to Clusters A and B, whereas "Neutral (Unique)" indicates both an equal distance and a code unique to Namgung Seonsaeng-jeon among the five works.

군집 중 유일하게 비극적 사건·도주(S12=a), 이인과의 만남(S15=a), 수련 과정(S16=a), 신이한 에피소드(S17=a)를 포함하여 B군집에 근접한 요소를 지닌다.

4-2 남궁선생전의 전환형 서사 구조: 수치 기반 심화 분석

표 3의 거리 행렬에서 남궁선생전과 A군집 사이의 평균 거리는 1.500(엄처사전 1.636, 손곡산인전 1.364의 평균), B군집과의 평균 거리는 2.182(장산인전 1.818, 장생전 2.545의 평균)이다. 두 값의 차이(0.682)는 A군집 내 엄처사전-손곡산인전 간 거리(0.636)에 근접한 매우 작은 수치로, 남궁선생전이 A군집에 속하면서도 B군집과의 거리가 군집 내 거리에 비해 그리 크지 않은 경계적 위치를 점유함을 보여 준다.

표 5는 22개 서사 단락 중, 5편 모두 동일한 코드를 지니 군집 변별력이 없는 3개 SID(S09·S13·S14)를 제외한 19개 SID를 대상으로, 남궁선생전의 코드값이 A군집(엄처사전·손곡산인전)과 B군집(장산인전·장생전)의 코드값에 각각 얼마나 가까운지 평균 거리를 산출하여 귀속 유형을 판정한 것이다. A군집과의 거리는 엄처사전·손곡산인전 각각의 코드값과의 거리를 평균한 값이며, B군집과의 거리는 장산인전·장생전 각각의 코드값과의 거리를 평균한 값이다.

표 5에서 22개 SID는 세 유형으로 판정된다. A형 귀속 9개(S01·S02·S03·S04·S05·S10·S18·S21·S22), B형 귀속 6개(S06·S08·S11·S15·S16·S17), 중립·독자 4개(S07·S12·S19·S20)이며, 분석에서 제외된 3개 SID(S09·S13·S14)는 5편 모두 동일한 코드를 지니 군집 변별력이 없다.

B형 귀속 6개 단락을 세분화하면, S15(이인과의 만남)·S16(수련 과정)·S17(신이한 에피소드)는 남궁두가 비극적 사건(S12) 이후 도교적 세계로 진입하는 전환 이후에 등장하는 단락이다. 남궁선생전에서 장로(長老)는 남궁두에게 혼과 정신을 단련하여 신선으로 탈바꿈하고 싶어하는 사람에게 있어 서야 더 말할 게 있겠냐고 이르는데, 이는 수련(S16)이 단순한 기예 습득이 아니라 신선화(神仙化)를 지향하는 존재론적 전환임을 보여 준다. S06(예술·잡기 능력)·S08(신분적 결핍)·S11(효행 서사)은 전환 이전 남궁두의 사회적 위치와 관련된 단락으로 B군집 패턴(없음=X)을 따른다. 이처럼 B형 귀속 단락은 인물의 생애 전반부와 후반부에 걸쳐 분산된 것이 아니라, 전환점 이전·이후에 기능적으로 구분되어 배치되어 있다.

중립·독자 유형 4개 중 S07(사상적 지향=c, 도교전환)은 남궁선생전에만 등장하는 독자 코드값으로, A군집의 a(유학)·B군집의 d(도교, 처음부터)와는 구별되는 전환형 서사 고유의 사상적 궤적을 표상한다. 남궁두는 유학에서 출발하여(S03=a: 유학적 품성) 도교로 전환하는(S07=c) 서사 경로를 거치는 유일한 인물인 것이다.

4-3 B군집: 신이·탈속형 이인 서사와 임진왜란의 신이성 검증

장산인전·장생전은 주인공 내력 불명(S01=b), 방술·신이 능력 보유(S05=a), 신이한 에피소드(S17=a), 임진왜란 관련 서사(S18=a), 사후 현현(S21=a)을 공유하며, 허균의 논란(S22)이 없다. 장산인전은 그의 부친이 상륙(商陸)을 먹고서 귀신을 보고 부릴 수 있었다고 서술하여, 신이한 능력이 가계로부터 전승되는 것으로 그린다. 이는 두 인물이 허균의 주관적 평가나 사상적 동류 의식의 대상이 아니라, 신이한 사실(事實) 자체를 기록 보존하는 방식으로 서술되었음을 함의한다.

두 편 사이에서도 차이가 있다. 장산인전은 이인과의 만남(S15=a)과 수련 과정(S16=a)을 포함하여 수련을 통한 도술 체득을 강조하는 반면, 장생전은 수련 없이 선전적·자연발생적 신이함을 지닌 인물이다. 또한 장생전에서만 예술·잡기 능력(S06=a2: 음악·모방)과 허균과의 직접 친교(S19=a)가 나타난다. 이는 장생이 신이한 방외인이면서도 허균이 젊은 시절 실제로 교유한 인물이라는 작품 내 특수성을 반영한다.

임진왜란 관련 서사(S18)는 B군집 2편에만 공통으로 나타나며, S18이 없는 A군집과의 거리는 4.0으로 최대치이다. 이는 임진왜란이라는 역사적 사건이 두 인물의 신이한 능력을 공식적으로 검증하는 장치로 서사에 편입되었음을 수치로 확인시킨다. 결말 유형(S20) 역시 두 군집이 완전히 분기한다. A군집의 결말은 자연사(a) 또는 불명확(b)인 반면, B군집의 결말은 신이한 죽음(a1: 장산인전)·신선화(a2: 장생전)로 처음부터 탈속적 삶의 종결을 예비한다.

4-4 S19·S22의 비대칭 분포와 허균 전 집필의 목적론

허균과의 직접 교류(S19=a)는 장생전에만 나타나고, 허균의 논란(S22=a)은 A군집 3편에만 존재한다. 두 단락의 비대칭적 분포는 허균의 5편 집필 동기가 단일하지 않음을 시사한다. 논란이 존재하는 세 편의 경우, 허균은 인물에 대해 적극적으로 평가를 내리고 자신의 사상적 공감을 기록하는 방식으로 집필에 임한다. A군집 인물(엄충정·이달·남궁두)은 허균이 이상적 삶의 전범으로 설정한 인물형이며, 논란은 그 평가 행위의 문학적 표현이다.

반면 직접 교류(S19)가 나타나는 장생전은 허균 자신의 젊은 시절 친교를 기록한 것으로, 개인적 회상(回想)의 성격이 강하다. 그러나 이 편에는 허균의 논란(S22)이 없다. 이는 장생이라는 인물이 허균의 사상적 동류로 평가하기 어렵거나, 신이한 사실의 기록이라는 목적이 사상적 평가보다 앞서는 경우임을 암시한다. S19·S22의 비대칭은 허균 전 문학이 "사상적 평가·이상화"와 "신이한 사실의 목적·보존" 두 가지 집필 목적 사이에서 분화되어 있음을 구조적으로 드러낸다.

V. 성소부부고 산문 인물 정보 추출의 예비 결과

본 연구의 직접 분석 대상은 허균의 5편 전(傳)이지만, 분석 방법의 확장 가능성을 검토하기 위하여 성소부부고 제4권부터 제26권까지의 산문 자료를 대상으로 인물 정보 추출을 예비적으로 수행하였다. 한국고전종합DB에서 제공하는 원문 및 번역문 자료를 문서 단위로 수집하고, 인명 표기와 한자 병기, 제목 수신자 및 주석의 인물 설명 형식을 중심으로 규칙 기반 후보 추출을 적용하였다.

표 6. 성소부부고 제4권~제26권 산문 인물 후보 추출 현황

Table 6. Candidate character extraction status for the prose in volumes 4-26 of Seongso Bubugo

Category	Results
Scope of Analysis	Prose in Volumes 4-26 of Seongso Bubugo
Number of Documents Analyzed	369 documents
Text Size Analyzed (characters)	604,579 characters
Number of Documents Containing Candidate Characters	353 documents
Number of Candidate Character Mentions	4,176 mentions
Number of Normalized Candidate Character Names	2,203 normalized names

Note. Person candidates are extracted based on Chinese character used together, name expression, and context related to names, letters, and letters. Expressions similar to the names of the same name, names, official names, and human names were left for follow-up review.

표 7. 인물 후보의 출현 위치별 분포

Table 7. Distribution of candidate character occurrences by location

Occurrence Location	Number of Mentions	Percentage (%)
Title	307	7.4
Main Text	3,259	78.0
Notes	610	14.6
Total	4,176	100.0

그 결과 총 369편, 604,579자 규모의 산문 자료에서 4,176건의 인물 후보 언급과 2,203종의 표준화 인물 표기를 확보하였다. 인물 후보의 출현 위치는 본문 3,259건, 제목 307건, 주석 610건으로 나타났다. 이는 허균 산문에서 인물 정보가 본문 서술에 집중되어 있으나, 제목의 수신자 표기와 주석의 인물 해설 역시 별도의 분석 단위로 활용될 수 있음을 보여 준다.

다만 본 예비 추출 결과에는 호·자·관직명·집단명 등 인명과 인접한 표현이 일부 포함될 수 있으므로, 본 연구에서는 이를 완결된 인물 데이터베이스가 아니라 후속 검증을 위한 후보 자료로 한정한다. 향후에는 인물별 동명 이인 판별, 국

적·신분·성별 등의 속성 정제, 허균과의 직접 교유 및 서신 관계 판별을 추가하여 성소부부고 산문 전체의 인물 표상과 관계 구조를 분석할 수 있을 것이다.

VI. 결 론

본 연구는 허균의 5편 전을 대상으로 22개 서사 단락을 설계하고 내용 유형을 인코딩하여, 해밍 거리 행렬 산출과 계층적 군집 분석을 수행한 시론이다. 분석 결과, 5편의 전은 A군집(유학-현실형: 엄처사전·손곡산인전·남궁선생전)과 B군집(신이-탈속형: 장산인전·장생전)으로 뚜렷이 분류된다. 이는 허균이 동일한 이인 서사 장르 안에서 유학적 교양 기반 인물과 도교·신이적 인물을 의식적으로 변별하여 서술하였음을 수치로 실증한 결과이다.

세부 분석에서 다음 다섯 가지 시사점이 도출되었다. 첫째, 엄처사전·손곡산인전의 거리 0.636은 5편 중 최솟값으로, 두 편이 처사형 서사 문법이라는 공통 구조 위에 있음을 증명한다. 둘째, 남궁선생전은 22개 SID 중 9개가 A형, 6개가 B형, 4개가 중립·독자로 분류되는 전환형 서사 구조를 지닌다. A군집 평균 거리(1.500)와 B군집 평균 거리(2.182)의 차이(0.682)는 군집 내 최소 거리(0.636)에 근접한 수치로, 남궁선생전이 A군집 경계에 위치함을 보여 준다. 셋째, 임진왜란 서사(S18)가 B군집에만 나타나며 A군집과의 거리는 최대(4.0)로, 임진왜란이 B군집 인물의 신이성을 공식 검증하는 서사 장치임이 수치로 확인된다. 넷째, 결말 유형(S20)은 A·B군집이 완전히 분기하여 장르 내 이분법적 서술 관습이 결말까지 일관됨을 보여 준다. 다섯째, S19·S22의 비대칭 분포는 허균 전 집필이 "사상적 평가·이상화"와 "신이한 사실의 목격·보존"이라는 두 목적 사이에서 분화되어 있음을 시사한다.

방법론적으로, 이본이 단일한 경우에도 복수의 작품을 비교 단위로 설정하면 해밍 거리 기반 서사 구조 비교가 유효하게 적용될 수 있음을 보여 준다. 다만 분석 대상이 5편에 한정되어 결과를 일반화하는 데 한계가 있으며, 서사 단락 분절과 인코딩 과정에서 연구자의 판단이 개입될 수 있다는 점을 유의해야 한다. 향후에는 본 연구에서 예비적으로 구축한 산문 인물 후보 자료를 정제하여, 성소부부고 문부 전체의 인물 출현 양상, 장르별 인물 표상, 허균과 인물 간 관계 구조를 분석할 필요가 있다.

참고문헌

- [1] J. S. Choi, "Heo Kyun's Narrative of Reality Politics and Practical Concerns - Focused on the <Namgung Seonsaeng Jeon> -," *Donam Ohmunhak*, Vol. 33, pp. 209-238, 2018. <https://doi.org/10.17056/donam.2018.33..209>

- [2] C. A. Kim, "The Imagery of a Dream in Heo, Gyun's Literature and the Meaning -Focused on Seongsobubugo," *Eomunyeongu*, Vol. 78, pp. 129-152, 2013. <https://doi.org/10.17297/rsll.2013.78..006>
- [3] Y. M. Shin, "Heo Gyun's Fate Acceptance and Record of 「Haemyeongmun」," *Journal of Korean Classical Chinese literature*, Vol. 43, No. 1, pp. 353-385, 2021. <https://doi.org/10.18213/jkcl.2021.43.1.012>
- [4] K. H. Shim, "Hö Kyun and Wang Shizhen," *Yeonmin Hak Chi*, Vol. 40, pp. 23-146, 2023. <https://doi.org/10.22996/yonmin.40..202308.23>
- [5] G. U. Lee, Heo Gyun's Theory of Writing and Its Realization, Ph.D. Dissertation, Sungshin Women's University, Seoul, 2010.
- [6] Y. Y. Kim, "Hö Gyun's Network of Social Relations as Observed Through His Correspondence" *Yeonmin Hak Chi*, Vol. 41, pp. 221-254, 2024. <https://doi.org/10.22996/yonmin.41..202402.221>
- [7] E. S. Yang, "Buddhistic View of Gyosan" *Gangwon Minsokhak*, Vol. 30, pp. 281-309, 2018.
- [8] M.-J. Lim, "A Study on the Ebon of 『Seongsobubugo』," *The Journal of Language & Literature*, Vol. 82, pp. 215-254, 2020. <https://doi.org/10.15565/jll.2020.06.82.215>
- [9] W. K. Kang and B. R. Kim, "A Study on the Transformation of Different Versions of the <So Hyeon-seong nok> Series by Computer - Focused on Ewha Womans University's 15-Volume Version and Kyujanggak's 21-Volume Version -," *Korean Language and Literature in International Context*, No. 80, pp. 115-135, 2019. <https://doi.org/10.31147/IALL.80.5>
- [10] W. K. Kang and B. R. Kim, "A Study on the Methodology of Digital Emotion Analysis for Classical Novels -For the Cloud Dream of the Nine," *DONG ASIA KODAEHAK(The East Asian Ancient Studies)*, No. 56, pp. 349-377, 2019. <https://doi.org/10.17070/aeaas.2019.12.56.349>
- [11] W. Choi and D. Kim, "Researches on Classifying Versions of Sipjangga by Measuring Similarities of Lexical Elements and using Hierarchical Clustering," *Journal of Korean Institute of Information Technology*, Vol. 10, No. 5, pp. 133-138, 2012.
- [12] W. Choi and D. Kim, "A Computational Approach to the Classification and Clustering of Tokkijeon through Pairwise Comparison of its Narrative Elements," *The Studies of Korean Literature*, No. 58, pp. 123-154, 2018. <https://doi.org/10.20864/skl.2018.04.58.123>
- [13] K. S. Kwon and D. Kim, "The Classification of "Sim Cheong-Jeon" Through a Computer Analysis Technique of Bibliographies," *Journal of Pansori*, No. 47, pp. 167-205, 2019. <https://doi.org/10.18102/jp.2019.04.47.167>
- [14] D.-K. Kim and H.-Y. Jeong, "A Study of Computational Literature Analysis Based Classification for a Pairwise Comparison by Contents Similarity in a Section of Tokkijeon, 'Fish Tribe Conference', *The Journal of the Korea Contents Association*, Vol. 22, No. 5, pp. 15-25, 2022. <https://doi.org/10.5392/JKCA.2022.22.05.015>
- [15] H. Kim, B. R. Kim, and Y. S. Lim, *Digital Humanities*, Seoul: HUFS Press, 2016.
- [16] I. T. Ryu, J. E. Gwak, K. S. Kwon, B. Kim, B. J. Kim, J. S. Kim, ... and H. Y. Han, *Digital Read and Write as Data - A Search for Digital Korean Literature*, Seoul: Bakmunsa, 2023.
- [17] E. H. Jo and D. H. Shin, "Analyzing Domestic Research Trends in Digital Literacy for the Elderly Using Text Mining," *Journal of Digital Contents Society*, Vol. 26, No. 2, pp. 381-392, 2025. <https://doi.org/10.9728/dcs.2025.26.2.381>

김해관(Hae-Kwan Kim)



1993년 2월 : 동의대학교

국어국문학과(문학사)

2002년 8월 : 동의대학교 교육대학원

국어교육(문학석사)

2024년 3월 ~ 현재 : 경희대학교 국어국문학과 박사과정 수료
※ 관심분야 : 고전산문, 허균, 디지털 인문학, 컴퓨터 문헌 분석 등

정화영(Hwa-Young Jeong)



1994년 : 경희대학교 전자계산공학과
(공학석사)

2004년 : 경희대학교 컴퓨터공학과
(공학박사)

1994년 ~ 1998년 : ㈜아주시스템 기술연구소 전임연구원
1998년 ~ 1999년 : ㈜CNA Research 기술연구소 전임연구원
2000년 ~ 2004년 : 예원대학교 만화계임영상학부 조교수
2005년 ~ 현재 : 경희대학교 후마니타스칼리지 교수
※ 관심분야 : 정보보호(Personal Information), 인공지능, 빅데이터 시스템 및 분석, 교육공학 등



김윤호(Yoon-Ho Kim)

1986년 : 경희대학교 전자공학과
(공학석사)

1992년 : 청주대학교 전자공학과
(공학박사)

2004년~2005년: Univ. of Auckland, NZ. CITR Lab.
Research Fellow

2012년~2014년: (사)사회안전학회 회장

2017년~2018년: (사)한국정보기술학회 회장

2021년~2022년: (사)한국디지털콘텐츠학회 회장

1992년~현 재: 목원대학교 컴퓨터융합학부 교수,
입학취업처장(이전), 교학부총장(이전),
감사실장(이전)

2008년~현 재: ISO/TC 292(Security & Resilience) Korea
Delegate

※관심분야 : Image Processing, Neuro Fuzzy System, IT
기반 사회안전망설계 등



안영훈(Young-Hoon An)

1991년 2월 : 경희대학교 국어국문학과
(문학사)

1993년 2월 : 경희대학교 국어국문학과
(문학석사)

1998년 8월 : 경희대학교 국어국문학과
(문학박사-고전문학전공)

1999년~2004년: 동의대학교 한국어문학부 교수

2018년~2019년: 동아시아고대학회 회장(현 평의원)

2023년~2026년: 우리문화회 회장

2005년~현 재: 경희대학교 국어국문학과 교수,
한국문화연구소 소장

※관심분야 : 고전문학 활용과 기획, 한문 번역 및 데이터
처리, 디지털 인문학, 컴퓨터 문헌 분석 등