

생성형 AI를 활용한 구비설화 자료의 구조화 방안 연구*

-『한국구비문학대계』 디지털 아카이브 자료를 중심으로-

한유진**

〈차례〉

1. 서론
2. 설화 자료 구조화의 이론적 기반
3. AI를 활용한 설화 자료의 예비 정리와 사건 단위의 설정
4. <해와 달이 된 오누이>의 구조화 적용과 그 의의
5. 결론

〈국문초록〉

본고는 <한국구비문학대계> 디지털 아카이브에 수록된 구비설화 자료를 생성형 AI 활용 환경에서 보다 안정적으로 식별·비교·재배열할 수 있도록 구조화하는 방안을 제안하였다. 이를 위해 <해와 달이 된 오누이> 유형 설화를 대상으로 삼아 AI를 활용하여 관련 자료를 탐색하고, 연구자가 이를 검토한 뒤 45편의 분석 대상을 확정하였다. 대상으로 삼은 45편에 대해 AI가 제시한 예비 구조화 결과를 검토하여 해석적 명명이나 층위가 다른 항목을 조정하고 관찰 가능한 행위와 상황을 중심으로 하는 '사건 단위' 설정 기준을 마련하였다.

이를 바탕으로 어머니의 외출, 호랑이와의 조우, 호랑이의 요구, 어머니의 죽음, 호랑이의 침입 시도, 어머니 여부 확인, 호랑이의 침입, 남매의 탈출, 남매의 피신, 호랑이의 추격, 남매의 응답과 속임, 남매의 기원, 남매의 상수, 해와 달로의 변신, 호랑이의 상수 시도, 호랑이의 추락, 수숫대의 유래에 이르는 17개의 사건 단위로 정리하였다. 이러한 기준을 각편 45편에 적용하고 공통적으로 반복되는 사건은 동일

* 본고의 논의를 보완하는 데 익명의 심사위원 세 분께서 제시해주신 조언이 큰 도움이 되었다. 지면을 빌려 깊이 감사드린다.

** 국민대학교 교양대학 조교수

하게 표시하되 각편별 변주 지점은 따로 기입하였다. 그 결과 사건 단위 기반 구조화는 표제가 다른 자료들을 사건 단위 기준으로 다시 식별할 수 있음을 확인하였다. 또한 각편의 변주를 특정 사건 단위가 실현되는 방식의 차이로 비교할 수 있었으며, 다양한 범주에 따라 자료를 다시 배열할 수 있는 가능성을 확인하였다. 다만 AI 구조화 결과에는 누락 등의 오류가 있을 수 있으므로 원문 대조와 연구자의 보정 과정이 필수적이다.

본고는 생성형 AI를 활용하여 구비설화 자료를 비교 및 추적 가능한 연구 데이터로 전환하는 방법론적 가능성을 제시했다는 점에서 의미가 있다. 이는 AI의 예비 처리와 인간 연구자의 검토 및 재설정이 결합된 자료 처리 방식(Human-in-the-Loop)으로, 이를 통해 <한국구비문학대계>와 같은 대규모 구비설화 자료가 단순한 원문 아카이브를 넘어 유형별 사건 구조와 변주 양상을 비교할 수 있는 지식 기반으로 확장될 수 있을 것이다. 나아가 본고의 사건 단위 스키마와 인간-AI 협업 기반 구조화 파이프라인은 다른 방대한 설화 유형으로 체계적으로 확장될 수 있을 뿐 아니라 국제 비교문학 연구에도 활용될 수 있는 가능성을 보여준다.

주제어 한국구비문학대계 디지털 아카이브, 구비설화, 생성형 AI, 구조화, 해와 달이 된 오누이, 사건 단위 스키마, 인간-AI 협업 기반 구조화, 디지털 인문학

1. 서론

본고는 한국구비문학대계 디지털 아카이브¹⁾(이하 <한국구비문학대계>²⁾에 수록되어 있는 구비설화 자료를 오늘날의 인공지능 활용 환경에서 보다 안정적으로 식별·탐색·비교·재배열할 수 있도록 구조화하는 방안을 제시하는 데 목적이 있다. <한국구비문학대계>에 DB화된 『한국구비

1) 한국구비문학대계 디지털 아카이브, <https://kdp.aks.ac.kr/inde/gubi>, 접속일 2026. 3.1.

2) 이후 서술부터는 출간된 『한국구비문학대계』와 구별하기 위해 디지털 아카이브는 <한국구비문학대계>로 표현하기로 한다.

문학대계』와 『증편 한국구비문학대계』는 전국의 구비설화를 총망라한 자료집으로, 출간된 구비설화 자료집 가운데 가장 방대한 규모를 지니거나 다수의 각편을 집적하고 있다는 점에서 구비설화 연구의 핵심 자료라 할 수 있다. 더욱이 이들 자료는 상당히 이른 시기부터 디지털화되어 공개됨으로써³⁾ 인쇄본으로만 존재하는 다른 구비설화 자료에 비해 검색 및 열람 등이 훨씬 용이하다.

그러나 그동안 이러한 접근성의 향상이 자료 활용의 효율로 이어진 것은 아니었다. 이는 <한국구비문학대계> 자료가 음성 녹취본과 이를 전사한 구술 채록본으로 제공되고 있기 때문이다. <한국구비문학대계>는 당시의 현장성을 포착한 원문 자료라는 점에서 귀중한 자료적 가치를 지니지만, 방언으로 표기된 전사 텍스트, 유형을 가늠하기 어려운 각편 제목들은 이 용자들에게 원활한 활용의 걸림돌로 작용했던 것이다.⁴⁾ 하지만 이러한 제약은 오늘날의 AI 환경에서 일정 부분 완화되었다. AI는 방언으로 이루어진 원문 설화 자료를 표준어로 변환하는 것은 물론이거니와⁵⁾ 특정 독자층의 수준에 맞게 재서술 및 요약 제공도 가능하다.⁶⁾ 또한 특정 인물, 사건,

3) 한국학중앙연구원에서는 1999년부터 왕실도서관 장서각에 소장된 자료와 한국학중앙연구원이 편찬 간행한 한국학 관련 연구 성과를 기반으로 한국학 데이터베이스 '왕실도서관 장서각 디지털 아카이브'를 구축하였다. 박현숙, 「설화 아카이브의 현황 분석과 활용 방안」, 『구비문학연구』 55, 한국구비문학학회, 2019, 214~215쪽.

4) 이는 구비설화 전승 양상의 지형을 거시적으로 파악하고자 하는 디지털 연구에서도 한계 요인으로 작용해온 것이다. 이러한 한계에 대해서는 김정은도 지적한 바 있다. 김정은은 단어 중심의 텍스트마이닝의 경우 누락되는 자료가 많고 주제적 접근에서는 관계성이 적은 자료까지 유의미한 단위로 포함되는 한계를 지적하면서 설화의 구조와 변이를 한눈에 알아보기 위한 데이터베이스 구축의 필요성을 제기한 바 있다. 김정은, 「『구비문학연구』 30년, 민담분석의 성과와 과제」, 『구비문학연구』 70, 한국구비문학학회, 2023, 126~127쪽 참고.

5) 표준어 변환에 대한 AI의 수행 수준은 다음 논문에서 구체적으로 확인할 수 있다. 한유진, 「한국 고전문학의 AI 번역, 그 가능성과 한계-〈아랑〉 설화 번역의 문화적 굴절을 중심으로-」, 『은지논총』 84, 한국은지학회, 2025, 257~260쪽.

6) 예컨대 한국어 학습자의 수준이나 초등학교, 중학생 등의 독자에 따라 동일한 자료를

모티프, 대강의 서사 전개를 단서로 일부이나마 해당 이야기 유형 및 각편을 찾아내고, 각편별 변주된 부분을 비교하는 일도 일정 수준 수행할 수 있다. 이는 최근 거대언어모델(Large Language Model)이 질문에 포함된 표현뿐 아니라 그 표현들 사이의 문맥적 관계와 의미적 연관성까지 고려하여 관련 자료를 탐색하기 때문이다.⁷⁾ 다만 이러한 생성형 AI는 문맥 추론은 가능한 반면 본질적으로는 학습된 언어 패턴에 기반하여 다음 단어를 확률적으로 생성하는 모델이기 때문에, 그 산출물을 안정적이고 일관된 자료 처리 결과로 무조건 수용하기는 어렵다.⁸⁾

이처럼 자료를 탐색하는 차원에서 그 결과가 완전히 정확하다고는 할 수 없지만 이제는 이용자가 <한국구비문학대계>로부터 필요한 자료를 어느 정도 확보하는 일은 이전보다 한결 수월해졌다고 할 수 있다. 다만 이러한 결과는 사용자의 질문 방식과 표현, 과업의 복잡성, 해석의 개입 정도에 따라 달라질 수 있으므로 그것이 곧 동일한 기준 아래 안정적이고 재현 가능한 자료 처리를 담보하는 것은 아니다. 실제로 AI의 응답은 프롬프트의 구성 방식에 따라 크게 달라질 수 있으며 동일한 질문에도 산출 결과가 조금씩 변동되는 비결정론적 특성을 지닌다.⁹⁾ 또한 이른바 환각(Hallucination)이라 불리는 사실과 다른 내용을 그럴듯하게 산출하는 문제 역시 여전히 남아 있다.¹⁰⁾ 이에 따라 현 단계에서 디지털 인문학이 고민해야 할 과제는 이러한 AI의 기술적 불안정성을 통제하고 비정형 자료들을 동일한 기준 아래 식별·비교·재배열할 수 있도록 하는 구조를 마련하는 일일 것이다.

본고는 이를 위해 <한국구비문학대계>의 원문 자료와 다양한 활용 방

서로 다르게 풀어쓰는 것도 가능하다.

7) 김성우, 『인공지능은 나의 읽기-쓰기를 어떻게 바꿀까』[전자책], 유투, 2024, 2장 참고.

8) 박태웅, 『박태웅의 AI 강의 2025』[전자책], 한빛비즈, 2024, 20% 참고.

9) 박태웅(2024), 위의 책, 21% 참고.

10) 박태웅(2024), 위의 책, 21~23% 참고.

식 사이를 매개할 수 있는 중간 층위를 설계하고자 한다. 이러한 중간 층위는 이용자의 자연어 질문과 설화 원문 사이를 연결해주는 번역 장치로 기능함으로써 구비설화 데이터를 기계 가독성을 갖춘 정형 데이터로 전환하는 데에 기여한다. 물론 이용자가 질문을 보다 정교화하고 다각도로 질의함으로써 AI 활용의 정확성을 일정 수준 높일 수는 있다.¹¹⁾ 그러나 이러한 방식은 개별 이용자의 질문 구성 능력에 기대는 임시적인 조정에 가깝다. 즉 질문이 달라질 때마다 결과가 달라질 수 있다는 점에서 이는 반복 가능한 자료 처리 기준이 되기 어려운 것이다. 이에 본고는 이러한 조정을 자료 내부의 구조화 기준으로 전환함으로써 질문의 표현이 달라지더라도 자료의 차원에서 일관되게 처리될 수 있도록 해결 방안을 탐색하고자 하는 것이다. 이는 인간 연구자의 자료 판별 기준과 분석 경험을 디지털 환경에서 즉각적으로 연산·비교·확장 가능한 시스템적 ‘스키마(schema)’로 정립하는 작업이며, 향후 대규모 아카이브의 자동화된 색인 및 검색 최적화를 위한 필수적인 토대를 마련하는 의미가 있다.

이러한 문제의식을 바탕으로 본고는 다음과 같은 방향으로 논의를 진행하고자 한다. 2장에서는 설화 자료 구조화의 이론적 기반을 검토하고 3장에서는 <해와 달이 된 오누이> 유형을 대상으로 생성형 AI를 활용한 자료 탐색과 예비 구조화 과정을 분석함으로써 사건 단위 기준을 설정할 것이다. 4장에서는 이러한 기준을 <해와 달이 된 오누이> 설화 각편 45편에 적용하여 그 효용과 한계를 검토하고자 한다.

11) 실제로 이야기 검색 시 ‘가족 질서가 회복되는 이야기’보다 ‘가족이 헤어졌다가 다시 만나는 이야기’를 요청했을 때 결과가 더 정확하게 제시됨을 확인할 수 있다.

2. 설화 자료 구조화의 이론적 기반

구비설화의 구조를 파악하는 시도는 일찍부터 이루어졌다. 이는 설화 분석을 위한 기반 작업이거니와 유형 분류, 비교 연구, 나아가 창작적 활용을 위해서도 그 필요성이 꾸준히 제기되어 왔다. 또한 최근에는 디지털 환경에서 원활한 활용 방식을 탐색한 연구들도 제출되고 있다.¹²⁾ 본고 역시 이러한 선행 연구의 연장선에서 <한국구비문학대계> 설화 자료를 오늘날의 AI 활용에 적합한 방식으로 구조화하고자 하는 것이다. 이는 AI를 통해 원문 자료만으로도 일정 수준의 탐색이 가능해진 상황에서 그러한 활용을 더 일관되며 추적 가능한 방식으로 전환하기 위한 구조를 마련하고자 하는 것이다.

이러한 구조화를 위해 이론적 배경으로 참고할 수 있는 것은 대표적으로 블라디미르 프로프(V. Propp)와 던데스(A. Dundes)의 분석틀이다. 프로프는 러시아 마법담을 대상으로 설화의 개별 인물이나 사건의 외양보다는 그것들이 이야기 전개 속에서 수행하는 ‘기능(function)’에 주목하였다.¹³⁾ 그에 따르면 기능은 이야기의 진행과 관련하여 행위가 지니는 의미로, 그 배열은 일정한 질서를 이룬다. 이러한 프로프의 논의는 설화를 일정한 사

12) 대표적으로 김정희와 이승은의 논의가 있다. 김정희는 비교적 이른 시기에 서대석의 단위담을 색인화하여 데이터베이스로 축적하면 구조 유형을 파악할 수 있는 검색어로 활용할 수 있음을 논의하였다. 한편 이승은은 <한국구비문학대계>의 이용자가 자료를 능동적으로 분절·재구성·활용할 수 있는 플랫폼적 환경이 충분히 마련되지 못했다는 점을 지적하면서 이들 자료를 대상으로 한 플랫폼 프로토타입 구축을 제안한 바 있다. 김정희, 「담춘 선생의 서사 구조와 의미 분석 방법을 활용한 검색 시스템의 설계-건국신화와 구비설화 연구를 중심으로」, 『구비문학연구』 63, 한국구비문학회, 2021, 31~68쪽.; 이승은, 「고전서사 플랫폼 구축 시론-『한국구비문학대계』 디지털 플랫폼 프로토타입 제작과 과제-」, 『고전과 해석』 48, 고전문학한문학연구학회, 2026, 35~68쪽.

13) 프로프는 31개의 기능을 제시하였다. 이에 대한 구체적인 내용은 다음 책을 참고할 수 있다. 블라디미르 프로프, 『민담형태론』, 유영대 옮김, 새문사, 2009, 39~88쪽.

건 배열의 질서를 지닌 구조적 총체로 파악하게 하는 계기를 마련하였다.¹⁴⁾ 던데스는 이러한 문제의식을 계승하면서 개별 모티프(motif)보다 이것이 놓이는 구조적 자리에 주목하고 이를 ‘단락소(motifeme)’로 개념화하였다. 그는 설화는 이러한 단락소들의 연쇄로 이루어진 것이며, 개별 모티프는 단락소를 구체적으로 실현하는 요소로 논의하였다.¹⁵⁾ 이에 따르면 설화 구조를 파악한다는 것은 개별 모티프의 나열이 아닌 그것들이 어떠한 자리에서 어떠한 관계로 배열되는가를 확인하는 데에 있는 것이다.¹⁶⁾ 이러한 프로프와 던데스의 논의는 설화의 구조를 모티프의 차원이 아닌 사건 배열의 질서로 이해할 필요가 있음을 보여준다.

이들의 논의는 오늘날 AI를 활용한 설화 자료의 구조화에 있어서도 여전히 유효하다. AI가 방대한 원문들 사이에서 특정 인물, 사건, 모티프 등을 단서로 관련 자료를 탐색할 수 있는 것은 단순히 질문에 사용된 어휘의 표면적 일치에만 근거하기보다 그 어휘들이 형성하는 맥락적 연관성과 의미적 관계를 바탕으로 처리하기 때문이다.¹⁷⁾ 따라서 설화 자료를 구조화하는 문제에서 중요한 것은 개별 키워드를 추출하는 것이 아니라 서사 속 사건 배열의 질서를 어떠한 단위와 수준에서 설정할 것인가에 있다.

이에 본고는 구조화의 기준으로 프로프의 기능과 던데스의 단락소를 참

14) 블라디미르 프로프(2009), 위의 책, 13~145쪽.

15) 던데스는 모티프가 인물, 사건, 배경과 같이 서로 다른 층위의 요소들을 함께 포괄하고 있기 때문에 구조 연구의 최소 단위가 되기 어렵다고 보았다. 따라서 개별 모티프는 구조의 기본 단위라기보다는 특정한 단락소를 구체적으로 채우는 실현 형태로 이해되며, 동일한 구조적 자리를 서로 다른 모티프가 채울 수 있는 경우 이를 알로모티프(allomotif)의 관계로 설명하였다. Alan Dundes, “From Etic to Emic Units in the Structural Study of Folktales”, *Journal of American Folklore*, vol. 75, No. 296, 1962, pp.95~105.

16) Alan Dundes, “Structural Typology in North American Indian Folktales”, *The Study of Folklore*, ed. Alan Dundes, Englewood Cliffs, N.J. : Prentice-Hall, 1965, pp. 208-209.

17) 김성우(2024), 앞의 책, 2장 참고.

고하되, 이를 실제 설화 자료에 적용 가능하도록 ‘사건 단위’로 조정하여 사용하고자 한다.¹⁸⁾ 이때 ‘사건 단위’는 사건 전개를 이루는 기본 국면을 의미하는 것으로, 개별 어휘나 표면적 모티프의 출현 여부가 아니라 이야기의 흐름 속에서 하나의 단락을 이루며 다음 전개와 관계를 맺는 사건의 배열 단위를 가리킨다.¹⁹⁾ 이는 개별 사건이 서사에서 수행하는 역할에 주목한다는 점에서 프로프와 던데스의 논의와 맞닿아 있지만, 이들 논의가 서사 질서를 추출하기 위한 이론적 범주인 것에 비해 본고의 ‘사건 단위’는 실제 설화 자료를 AI 활용 환경에서 식별·비교·재배열하기 위해 설정한 자료 처리상의 작업 단위라는 점에서 차이가 있다. 즉 프로프의 기능이 행위가 이야기 전개 속에서 지니는 기능적 의미를 가리키고, 던데스의 단락소가 서로 다른 모티프들이 채울 수 있는 구조적 자리를 설정하는 개념이라면, 본고의 사건 단위는 설화 각편들에서 관찰 가능한 행위와 사건을 공통 항목으로 분절하여 각편별 변수의 출현 위치와 실현 방식을 대조하기 위한 ‘기준 축인 것이다. 따라서 사건 단위는 보편적 구조 모형을 추출하려는 추상적인 이론 범주라기보다 각편의 공통 사건과 변수 정보를 동일한 형식으로 기입하고 이를 검색·비교·재배열할 수 있게 하는 ‘자료 처리 스키마’라고 할 수 있다.

요컨대 <한국구비문학대계> 자료를 프로프의 기능이나 던데스의 단락소

18) 다만 본고는 프로프나 던데스의 분석을 재현하는 데에 있지 않고, 이들의 관점을 바탕으로 실제 자료의 식별·비교·재배열에 적합한 기준을 마련하는 데에 있다.

19) 본고의 사건 단위는 이야기의 흐름 속에서 하나의 국면을 이루고 다음 전개와 연결된다는 점에서 서대석의 논의한 ‘단위담’과 일정한 측면에서 유사하다. 그러나 단위담은 주체가 어떤 목적을 가지고 움직여서 그 목적을 달성하는 데 성공하거나 실패하여 이야기로서 ‘완결된 의미’를 담아내는 서사의 기본 단위라는 점에서 본고의 ‘사건 단위’와 다르다. 이때 ‘완결된 의미’는 심우장이 언급한 ‘서사적 불균형과 균형의 1회적 조합’으로 이해할 수 있다. 서대석, 『이야기의 의미와 해석』, 세창출판사, 2020, 43쪽; 심우장, 「설화에 나타난 서사 확장의 두 패턴」, 『구비문학연구』 65, 한국구비문학학회 2022, 119쪽.

에만 준하여 구조화할 경우 서사의 일반적인 틀을 파악하는 데에는 유효할 수 있으나, 실제 아카이브에 축적된 수많은 각편을 정밀하게 분절하고 비교 가능한 자료로 전환하기 위해서는 보다 구체적인 작업 단위로의 조정이 필요하다. 이에 따라 본고는 분석 단위를 데이터 처리에 맞는 구체적인 작업 층위로 조정하고, 비정형 구술 텍스트를 AI 환경에서 정형 데이터로 전환할 수 있는 사건 단위를 제안하고자 하는 것이다. 다음 장에서는 이러한 사건 단위 기준이 생성형 AI를 활용한 설화 자료의 예비 정리 과정에서 구체적으로 어떻게 도출되고 설정되는지 그 실제적 방안을 살펴보고자 한다.

3. AI를 활용한 설화 자료의 예비 정리와 사건 단위의 설정

이 장에서는 <해와 달이 된 오누이> 유형을 대상으로 설화 자료의 구조화 방안을 구체적으로 검토해보고자 한다. <해와 달이 된 오누이>를 분석 대상으로 삼은 이유는 이 유형은 다수의 각편을 보유하고 있어 구조화 기준을 적용하여 그 양상을 확인하기에 적절하기 때문이다. 또한 대부분의 각편이 ‘해와 달이 된 오누이’라는 공통된 제목으로 제시되어 있으면서도 ‘호랑이 이야기’, ‘팔죽 할머니’, ‘하늘로 올라간 남매’, ‘목할머니와 호랑이’, ‘아이 잡아먹으려는 호랑이’, ‘사람 잡아먹는 호랑이’ 등과 같이 서로 다른 제목들도 확인된다는 점에서, 제목의 표면적 일치 여부만으로는 포착되기 어려운 동일 유형의 구조화 양상을 파악하기에 적합하다고 판단했기 때문이다.

본고는 이 유형 설화를 대상으로 사건 단위를 설정하기 위하여 다음의 단계로 작업을 진행하였다. 먼저 ①**생성형 AI²⁰⁾를 활용하여 관련 자료를**

20) 본고에서 자료의 예비 탐색, 목록화, 사건 국면의 예비 정리에 활용한 생성형 AI는 OpenAI의 ChatGPT(GPT-5.4 Thinking)이다. OpenAI는 이 모델을 GPT-5 계열

탐색하고 목록화한 뒤 ②자료 누락 여부를 점검하였다. 다음으로 ③아카이브에서 수집한 각편 텍스트를 AI에게 제공한 뒤 사건 단위 수준의 예비 정리를 요청하였다. 이후 그 ④결과를 검토하여 구조화 기준을 설정하였다. 즉 본고에서 구조화 단위의 설정은 AI의 예비 정리와 연구자의 조정 및 재설정(인공지능)이 결합된 방식(Human-in-the-Loop)으로 수행된 것이다.

3.1. AI를 활용한 관련 자료의 탐색 및 목록화

먼저 생성형 AI의 웹 탐색 기능을 활용하여 <한국구비문학대계> 아카이브에 있는 <해와 달이 된 오누이> 자료를 탐색하도록 하였다. 이때 탐색은 자연어 질의를 통해 어머니가 호랑이에게 해를 입고 남매가 도망친 뒤 해와 달이 되는 서사를 간략하게 제시하는 방식으로 이루어졌다.

한국구비문학대계 <https://kdp.aks.ac.kr/inde/gubi> 디지털 아카이브야. 여기서 호랑이가 엽마를 잡아먹고, 아이들까지 잡아먹으려고 하자 아이들이 도망가서 해와 달이 된 이야기를 찾아줘.

해당 아카이브에서 바로 찾아보겠습니다. 제목이 꼭 「해와 달이 된 오누이」가 아닐 수도 있어서, 내용 기준으로 관련 각편을 먼저 추려 보겠습니다.

관련 각편은 여러 편이 보입니다. 지금은 질문하신 서사 요소와 가장 정확히 맞는 각편을 확인해서, 바로 열람할 수 있는 링크와 핵심 내용을 정리하겠습니다.

[그림 1] AI를 활용한 <한국구비문학대계> 아카이브 서사 기반 탐색

이에 따라 AI는 제목의 일치 여부가 아니라 내용 수준의 유사성을 기준으로 관련 각편을 추출하겠다고 응답하였다. 그러나 실제 산출된 결과는

의 reasoning model로 설명하며, 복잡한 실제 업무, 긴 맥락을 요구하는 과제, 도구 활용이 수반되는 작업에서 향상된 수행 능력을 보이는 모델로 소개하였다. OpenAI (2026.3.18.), “Model Release Notes”, https://help.openai.com/en/articles/9624314-model-release-notes?utm_source=chatgpt.com, 접속일 2026.5.17.

‘해와 달이 된 오누이’라는 제목을 가진 각편들이 주로 제시되었고, 제목이 다른 이야기들까지는 안정적으로 포괄하지는 못한 결과가 확인되었다.²¹⁾ 이에 본고는 프롬프트의 조건 규칙과 질의 표현을 단계적으로 조정하면서 자료의 재탐색을 반복하였고, 산출 결과물에 더 이상 새로운 각편이 추가되지 않는 상태에 이를 때까지 정제 작업을 수행하여 총 20편의 이야기를 확보할 수 있었다.²²⁾

이후 필자가 <한국구비문학대계> 아카이브에서 직접 자료를 검색한 결과와 AI를 통해 확보한 자료군을 대조하였다. 그 결과 ‘해와 달이 된 오누이’라는 표제만으로 검색되는 자료는 총 161편이 확인되었으며, 이를 다시 1980년대에 채록된 『한국구비문학대계』 자료로 한정하자 15편이 남았다.²³⁾ 그러나 실제 <한국구비문학대계> 아카이브에 수록된 1980년대 조

21) AI는 총 11편의 각편을 제시하며 다음과 같이 응답하였다. “아카이브에서 질문하신 서사와 같은 유형으로 확인되는 이본들을 우선 모아보면 아래와 같습니다. 기준은 ①어머니가 호랑이에게 잡아먹힘, ②호랑이가 어머니 행세를 하며 아이들에게 접근함, ③아이들이 나무/하늘로 피신함, ④해·달 또는 하늘 승천으로 마무리됩니다. 총 11편의 이야기가 결과로 산출되었는데, 검색상 드러난 자료를 기준으로 정리한 것이며, 제목이 완전히 다르거나 검색엔진에 잘 잡히지 않는 각편은 더 있을 수 있습니다.” 이때 제시된 자료 중 10편의 각편 제목이 ‘해와 달이 된 오누이’이며, 1편은 ‘오누이와 호랑이’였다. OpenAI, ChatGPT(GPT-5.4 Thinking, 활용일자 2024.4.15.), 대형언어모델(LLM), <https://chatgpt.com/>, 2026.

22) 본고는 총 여섯 차례의 질의 및 프롬프트 최적화 과정을 추가로 수행하였으며, 이에 대해 AI는 “검색어, 제목, 모티프군까지 나뉘서 누락되기 쉬운 자료까지 검색하여 최대한 다시 훑어본 결과, 문자 텍스트가 확인되는 같은 유형·근연 유형 자료는 현재 기준으로 아래까지 확인된다”고 응답하였다. AI는 이를 바탕으로 총 20편의 설화를 다음과 같은 형태로 목록화하여 제시하였다. “1. 해와 달이 된 오누이 — 경기 김포시 통진읍 웅정리, 2009.02.27., 권애기. 어머니가 메밀떡을 얻어 오다 호랑이에게 잡아먹히고, 아이들은 손을 확인한 뒤 도망쳐 나무로 올라가 하늘의 동아줄로 올라갑니다. 호랑이는 썩은 동아줄을 타다 수수깡에 떨어집니다.” https://kdp.aks.ac.kr/inde/gub_i?id=POKS.GUBI.GUBI.2_10814” OpenAI, ChatGPT(GPT-5.4 Thinking, 활용일자 2026.4.15.), 대형언어모델(LLM), <https://chatgpt.com/>, 2026.

23) <한국구비문학대계> 아카이브에는 『한국구비문학대계』와 『증편 한국구비문학대계』 자료가 수록되어 있다. 이때 『한국구비문학대계』 자료는 1979~1985년, 『증편 한국구

사된 <해와 달이 된 오누이> 이야기는 총 45편으로,²⁴⁾ AI가 식별한 자료는 이 가운데 8편에 불과했다. 즉 이러한 검색만으로는 상당수의 자료가 누락되었음이 확인된다. 이러한 대조 결과를 반영하여 본고에서 분석 대상으로 삼은 『한국구비문학대계』에 실린 <해와 달이 된 오누이> 자료 목록 45편을 정리하면 다음 <표 1>과 같다.²⁵⁾

<표 1> <해와 달이 된 오누이> 자료 목록

	각편 제목	제보자 정보	수록 위치(링크)
1	해와 달이 된 오누이	오수영(여,68)	1-9, 209~212
2	해와 달이 된 오누이	이옥녀(여,70)	2-7, 513~516
3	해와 달이 된 오누이	최귀성(여,43)	5-1, 49~52
4	해와 달이 된 오누이	김순이(여,81)	1-7, 272~275
5	해와 달	김현녀(여,86)	5-2, 529~532
6	수수대가 빨간 유래	김성례(여,76)	6-4, 881~883
7	해와 달이 된 오누이	황원수(여,73)	6-11, 613~616
8	해와 달이 된 오누이	심운택(남,60)	2-6, 473~476
9	해와 달이 된 오누이	정숙이(여,78)	1-5, 308~311
10	수수깡이 빨간 이유	박옥순(여,66)	1-7, 761~762
11	호랑이 피 묻은 수수대	김정금(여,52)	6-7, 72~74
12	수수대가 빨간 이유	임묘금(여,66)	6-8, 728~729
13	수수대가 빨간 유래	정연옥(여,87)	2-5, 104~107

비문학대계』는 2008~2018년에 조사가 이루어진 자료이다. 본고에서 이 두 자료집 가운데 『한국구비문학대계』 자료만을 분석 대상으로 한정된 이유는 『문학치료서사사전』에 정리된 각편 목록이 『한국구비문학대계』 수록 자료를 대상으로 하고 있어, 이를 바탕으로 AI 산출 결과와 대조·검토하기에 적합하다고 판단했기 때문이다.

24) 이는 『문학치료서사사전』에 정리된 각편과의 대조를 통해 확인한 결과이다. 이 책에 정리된 <해와 달이 된 오누이> 각편은 총 47편인데, 그 중 다음 두 편은 아카이브에서는 확인되지 않는 누락 자료로 판단된다. 문학치료 서사사전 연구모임, 『문학치료서사사전2』[전자책], 휴머니스트, 2002. <무지개가 된 형제>(『한국구비문학대계』 6-5, 388~390쪽), <호랑이와 오누이>(『한국구비문학대계』 7-1, 427~429쪽)

25) <표 1>에서 수록 위치는 해당 각편이 실린 『한국구비문학대계』의 책번호와 면수를 제시한 것으로, 해당 표기에는 아카이브 원문으로 이동할 수 있는 링크를 연결하였다.

14	해와 달의 유래	김선옥(여,52)	3-4, 782~788
15	해와 달이 된 남매	고정숙(여,60)	4-4, 330~333
16	해와 달이 된 오누이	이항훈(남,71)	1-4, 81~84
17	해와 달이 된 오누이와 호랑이	최정자(여,41)	5-3, 312~315
18	해와 달의 유래	이화옥(여,69)	3-2, 408~411
19	수수깡이 빨간 것은 호랑이 핏자국	정명자(여,44)	5-2, 516~519
20	별순이와 달순이와 수수대의 핏자국	이점순(여,54)	7-17, 559~561
21	수궁대이가 빨간 까닭	조만순(여,50)	7-5, 52~53
22	해와 달이 된 남매	이연순(여,55)	4-5, 751~752
23	팔죽 할머니 이야기	최순금(여,63)	7-12, 144~147
24	수수깡의 빨간 것은 호랑이의 핏자국	임영순(여,50)	5-2, 629~632
25	콩단이와 팥단이와 호랑이	백연식(여,73)	7-15, 481~483
26	수숫대가 붉은 이유	김오봉(남,78)	1-8, 321~323
27	해와 달이 된 오누이	김철중(남,28)	6-8, 22~23
28	수숫대가 빨간 이유	박석원(남,48)	6-10, 304~305
29	해와 달이 된 남매	문순덕(여,59)	6-12, 627~630
30	해와 달이 된 오누이	이명분(여,71)	7-8, 342~345
31	해와 달이 된 자매	김성련(여,87)	8-12, 350~352
32	하늘로 올라간 남매	박용애(여,66)	4-5, 178~179
33	묵할머니와 호랑이	정복란(여,72)	7-4, 118~120
34	팔죽 할머니	김태강(여,74)	7-10, 330~331
35	수숫대가 붉은 유래	김용화(남,62)	2-7, 123~125
36	수숫대가 빨강게 된 내력	안금옥(여,63)	7-10, 642~644
37	아이 잡아먹으려는 호랑이	최윤순(여,52)	7-6, 199~202
38	오누이와 호랑이	박을순(여,69)	2-7, 236~239
39	해와 달이 된 오누이	권일분(여,68)	7-8, 509~513
40	해와 달이 된 오누이	구또현(여,75)	8-4, 555~557
41	해와 달이 된 남매	선영자(여,47)	6-12, 585~587
42	호랑이 이야기	김희목(여,61)	1-2, 422~423
43	수숫대가 빨간 이유	김기남(남,61)	6-10, 73~74
44	사람 잡아 먹는 호랑이	윤추월(여,66)	9-1, 674~679
45	팔죽 할멈과 호랑이	안옥순(여,76)	2-6, 560~561

요컨대 이러한 결과는 생성형 AI가 관련 자료의 예비 탐색 도구로서는 활용될 수 있으나, 정확하고 안정적인 자료 확보를 위해서는 연구자의 검

토와 재선별 과정이 반드시 필요함을 보여준다. 동시에 생성형 AI의 산출 결과는 사용자가 자료 판별 기준을 점차 구체화해 갈수록 그 기준을 보다 정교하게 반영하는 방향으로 조정될 수 있음도 확인된다.

3.2. 원문 텍스트를 활용한 예비 정리와 사건 단위 설정의 기준

이 절에서는 앞서 검토를 마친 <표 1>의 45편 자료를 바탕으로 생성형 AI를 활용한 예비 구조화 결과를 검토하고 이를 통해 사건 단위 설정의 기준을 구체화하고자 한다. 이를 위해 원문 하이퍼링크와 채록 원문이 수록된 데이터세트 파일을 AI에게 제공하고 이를 기준으로 각편의 원문을 참조하여 사건 단위로 구조화하도록 요청하였다. 이때 요청한 작업은 각 이야기의 제목을 붙이고 사건 단위에 따라 서사를 분절한 뒤 그 구조화 방식을 함께 설명하도록 하는 것이었다. 그러자 AI는 입력된 데이터세트의 문맥적 관계와 의미적 연관성을 분석하여 45편 전체를 아우를 수 있는 공통 서사 축을 상위 수준에서 제시하고, 이를 기준으로 각편별 사건 배열을 예비적으로 분절·정리하여 정형화된 스프레드시트 규격의 구조화 데이터로 산출하였다.

호	자료명	외증/거기	노상 조우·회생	가장	정체 확인	탈출	피신	추격 교란	구원	결말	비고
1	해와 달이 된 오누이 (1-9, 209-212)	베/무명 매리 감 종말이 일하러 감	먹을거리 요구 / 신체 옷 요구 / 어머니 회생	어머니 가장 (목소리 변명)	순 확인 / 목소리 의심 / 깨우는 소리/먹는 것 질문	배변 핑계 / 문 빠져나감	나무 위 / 생/ 우물 근처	기름/소통 등 허위 정보/도끼/자귀 방법 제시	하늘에 기도 / 새 울/현 줄	해 달 됨	수순대 붉은 이 유 결부
2	해와 달이 된 오누이 (2-7, 513-516)	종말이 일하러 감	먹을거리 요구 / 신체 옷 요구 / 어머니 회생			배변 핑계	나무 위		하늘에 기도 / 새 울/현 줄	해 달 됨	수순대 붉은 이 유 결부
3	해와 달이 된 오누이 (5-1, 49-52)	종말이 일하러 감	먹을거리 요구 / 신체 옷 요구 / 어머니 회생	어머니 가장 (목소리 변명)	순 확인 / 목소리 의심 / 깨우는 소리/먹는 것 질문	문 빠져나감	나무 위 / 생/ 우물 근처	기름/소통 등 허위 정보/도끼/자귀 방법 제시	하늘에 기도 / 새 울/현 줄	해 달 됨	아기/늙은 회생 / 수순대 붉은 이 유 결부
4	해와 달이 된 오누이 (1-7, 272-275)	베/무명 매리 감	먹을거리 요구 / 신체 옷 요구 / 어머니 회생	어머니 가장 (목소리 변명)	순 확인 / 목소리 의심 / 깨우는 소리/먹는 것 질문	배변 핑계 / 문 빠져나감	나무 위 / 생/ 우물 근처	기름/소통 등 허위 정보/도끼/자귀 방법 제시	하늘에 기도 / 새 울/현 줄	해 달 됨	수순대 붉은 이 유 결부

[그림 2] 생성형 AI의 사건항 기반 예비 구조화 결과

이처럼 생성된 예비 구조화 결과는 45편의 각편을 공통축에 따라 정리함으로써 각편별 서사 전개의 면모를 어느 정도 파악할 수 있게 해주었는데, 이는 사건 단위 설정 기준을 조정하는 출발점이 되었다. 하지만 AI가 생성한 예비 구조화 결과의 항목들은 그 배열이 단일한 기준만으로 이루어지지 않아 재조정이 필요하다. 이를테면 ‘무명 매러 감’, ‘문을 열어줌’, ‘도망감’과 같이 서사에 나타난 구체적 행위를 기술한 항목과 ‘위장된 귀환’, ‘정체 확인’, ‘위험 인지’와 같이 사건의 국면을 요약한 항목, 그리고 ‘희생’이나 ‘구원’과 같은 해석적 용어가 사건 단위로 함께 뒤섞여 있는 것이다.²⁶⁾ 이처럼 행위적 차원, 국면의 차원, 해석의 차원이 함께 혼재되어 있을 경우 각 항목은 동일한 비교 단위가 될 수 없다.

이는 구조화된 출력(Structured Outputs) 제약 조건을 명시하지 않을 경우 거대언어모델(LLM)이 문맥을 추론하는 과정에서 데이터의 추상화 층위를 엄밀하게 판별하지 못하여 발생하는 기술적 한계이다. 특히 해석적 용어는 사건 단위의 명칭으로 사용하기에 부적합하다. 예컨대 ‘희생’은 타인이나 어떤 목적을 위해 자신을 내어놓는 의미이므로 어머니가 호랑이에게 죽임을 당하는 사건을 ‘희생’으로 규정하면 그 사건은 윤리적 의미를 갖는 장면으로 재맥락화되는 것이다.²⁷⁾ 마찬가지로 동아줄이 내려와 남매가 하늘로 올라가는 장면을 ‘구원’으로 명명하면 사건의 구체적 형태와 별개로 구출, 회복 등과 같이 장면이 지닌 기능과 의미를 확정해버리는 결과가 된다.

이처럼 해석이 개입된 용어는 분석 단계에서는 유의미한 구조화를 가능하게 할 수 있지만²⁸⁾ 본고의 구조화 작업 단위로서는 적절하지 않다. 이는

26) 제시된 [그림 2]는 AI가 제시한 스프레드시트 파일 가운데 일부만을 제시한 것이기 때문에 [그림 2]에서는 이러한 내용을 찾아볼 수는 없다.

27) 더욱이 이 이야기에서 어머니가 호랑이에게 죽임을 당하는 사건이 자식을 위한 선택이거나 이외 다른 목적 때문이 아니라는 점에서 서사적으로도 희생으로 해석되기 어렵다.

본고의 구조화 목적이 서로 다른 제목과 표현을 가진 자료들을 동일한 기준 아래 식별뿐 아니라 비교하고 재배열까지 가능하도록 하는 데 있기 때문이다. 즉 생성형 AI의 비결정론적·확률적 산출을 통제하고 정렬 가능한 데이터로 전환하기 위해서는 개별 이야기를 요약하기 위한 차원이 아니라 동일 유형의 자료를 반복적으로 대조할 수 있도록 하는 안정적인 시스템적 스키마로서의 사건 단위가 설정될 필요가 있다.

이처럼 예비 구조화 결과를 검토한 것을 바탕으로²⁹⁾ 인간 연구자의 전문적 보정을 거쳐 확정된 ‘사건’ 단위 기준을 설정하면 다음과 같다.

1. 사건명은 해석을 최대한 배제한 관찰 가능한 행위와 상황을 서술하는 방식으로 한다.
2. 하나의 사건 단위는 서사 진행에서 하나의 사건 국면을 이루는 행위로 설정한다.
3. 세부 표현의 차이는 사건 단위를 나누는 기준이 아니라 변주 사항으로 별도로 처리한다.
4. 사건명은 가능한 주체와 행위가 드러나는 서술형으로 통일하여 배열의 수준을 안정화한다.

이러한 기준을 적용할 때 동일 유형의 이야기는 제목의 차이와 세부 표

28) 다음 신동훈의 지적은 이를 잘 보여주는 연구이다. 신동훈은 <신바닥이> 설화를 일반적 관행에 따라 ‘결핍-해결의 시도-위기-극복’의 순차구조로 정리할 경우 그것이 너무 평면적이며 실질적인 서사적 의미해석으로 나아가기에 너무 추상적이라고 지적하면서 서사적 화두를 축으로 한 순차구조가 서사의 초점과 방향성을 담고 있으며 구체적인 서사적 의미를 포함한 형태로 재서술될 수 있다고 논의하였다. 신동훈, 「서사적 화두를 축으로 한 화소·구조 통합형 설화분석 방법 연구」, 『구비문학연구』 46, 한국구비문학학회, 2017, 64~65쪽 참고.

29) 이때 AI의 예비 구조화는 한 차례의 시도에서 얻어진 것이 아니라 필자가 사건의 개념과 명명 원칙을 조정하며 프롬프트를 반복적으로 정제하고 산출 결과를 재검토한 결과이다.

현의 변주에도 불구하고 비교 가능한 방식으로 구조화될 수 있을 것이다.

4. <해와 달이 된 오누이>의 구조화 적용과 그 의의

이 장에서는 앞서 정리한 사건 단위 기준을 실제 <해와 달이 된 오누이> 유형 설화에 적용한 결과를 검토하고, 이를 통해 사건 단위 기반 구조화가 어떠한 수준에서 작동하는지를 확인해보고자 한다. 이는 동일 유형의 각편들을 공통 사건 아래 배열해봄으로써 이들 자료가 서사의 공통 구조와 변주된 지점들을 비교 가능한 형식으로 드러낼 수 있는지를 파악하고 이를 바탕으로 구조화의 실질적 효용을 검토하고자 하는 것이다.

이를 위해 앞서 설정한 사건 단위의 원칙을 적용하여 <해와 달이 된 오누이>를 구조화하였으며, 그 결과는 다음과 같다.

1. 어머니가 외출한다.
2. 귀갓길에 호랑이와 조우한다.
3. 호랑이가 먹을거리나 신체 일부를 요구한다.
4. 어머니가 호랑이에게 잡아먹힌다.
5. 호랑이가 집으로 찾아와 남매에게 문을 열라고 한다.
6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다.
7. 호랑이가 어머니인 척하고 아이들이 문을 열어준다.
8. 남매는 호랑이 존재를 알아차리고 밖으로 탈출한다.
9. 남매가 나무 위로 올라가 숨는다.
10. 호랑이는 나무에 올라가는 방법을 묻는다.
11. 남매는 응답하거나 호랑이를 속인다.
12. 남매가 하늘에 빈다.
13. 줄이 내려와 남매가 하늘로 올라간다.

14. 남매는 해와 달이 된다.
15. 호랑이도 하늘에 빌어 줄을 타고 올라간다.
16. 줄이 끊어져 호랑이가 수수밭에 추락한다.
17. 호랑이의 피로 수숫대가 붉은 것이다.

위 사건항은 45편 전체를 동일한 기준 아래 배열하기 위해 반복적으로 나타나는 행위 국면을 정리하여 비교 단위로 설정한 것이다. 즉 어머니의 외출, 호랑이와의 조우, 호랑이의 요구, 어머니의 죽음, 호랑이의 침입 시도, 어머니 여부 확인, 호랑이의 침입, 남매의 탈출, 남매의 피신, 호랑이의 추격, 남매의 응답과 속임, 남매의 기원, 남매의 상승, 해와 달로의 변신, 호랑이의 상승 시도, 호랑이의 추락, 수숫대의 유래에 이르기까지 일련의 국면을 공통항으로 설정하였다. 이때 이러한 행위 국면은 특정 각편에만 나타나는 세부 표현을 구체적으로 읊진 것이 아니라 이들을 포괄할 수 있는 상위 수준에서 정리하였다. 예컨대 3항의 ‘먹을거리나 신체 일부’는 떡, 밥, 메밀묵, 범벅, 옷, 팔, 다리 등의 구체적 대상을 포함한 것이며, 11항에서 ‘남매는 응답하거나 호랑이를 속인다’는 남매의 나무 피신 이후 호랑이의 추격에 대한 대응을 하나의 국면으로 묶어낸 것이다.

이러한 17개의 사건 단위가 45편의 각편에 어떻게 배열되어 있는지는 표로 정리하였다. 이때 각편에서 공통적으로 반복되는 사건은 ○로 표시하고, 각편마다 다르게 제시되는 부분은 변주 사항으로 따로 정리하도록 하였다. 이러한 항목별 사건 기입 원칙은 [그림 3]으로 정리할 수 있으며, [그림 4]는 이를 적용한 구조화 결과의 일부를 제시한 것이다.

항목	기입 원칙
1. 어머니가 외출한다	공통적으로 확인되면 ◦, 외출 형태가 특이하면 간단히 변주 기입
2. 귀갓길에 호랑이와 조우한다	고개/재/길목에서 호랑이와 조우하면 ◦
3. 호랑이가 먹을거리나 신체 일부를 요구한다	요구 대상이 다르면 ◦ 대상만 간단히 기입
4. 어머니가 호랑이에게 잡아먹힌다	단순 포식은 ◦, 팔다리 절단 뒤 포식이면 변주 기입
5. 호랑이가 집으로 찾아와 남매에게 문을 열고도 한다	집에 와 문 열고도 하면 ◦
6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다	손/목소리/애기.젖 등 확인 표지를 기입
7. 호랑이가 어머니인 척하고 아이들이 문을 열어준다	문을 열어주면 ◦, 문을 부수고 들어오면 변주 기입
8. 남매는 호랑이 존재를 알아차리고 밖으로 탈출한다	탈출 방식이 특이하면 그 방식 기입
9. 남매가 나무 위로 숨는다	나무 종류가 드러나면 그 이름 기입
10. 호랑이는 올라오는 방법을 묻는다	올라오는 방법을 묻는 장면이 있으면 ◦
11. 아이들이 응답하거나 호랑이를 속인다	기름/도끼/자귀/소동 등 응답 내용을 기입
12. 남매가 하늘에 반다	하늘/하느님에게 비는 장면이 있으면 ◦
13. 줄이 내려와 남매가 하늘로 올라간다	줄이면 ◦, 방식이 함께 내려오면 줄 방식 기입
14. 남매는 해와 달이 된다	해·달 배정이나 교체 여부가 드러나면 기입
15. 호랑이도 하늘에 빌어 줄을 타고 올라간다	호랑이도 기도하거나 줄을 타면 ◦
16. 줄이 끊어져 호랑이가 수수대밭에 추락한다	수수밭/대밭 등 추락 지점을 간단히 기입
17. 호랑이의 피가 묻어 수수대가 붉은 것이다	수수대가 붉어진 유래가 설명되면 ◦

[그림 3] 사건 항목별 기입 원칙

번호	제목	2. 귀갓길에 어머니가 외출한다	3. 호랑이가 귀갓길에 호랑이와 조우한다	4. 어머니가 호랑이에게 잡아먹힌다	5. 호랑이가 집으로 찾아와 남매에게 문을 열고도 한다	6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다	7. 호랑이가 어머니인 척하고 아이들이 문을 열어준다	8. 남매는 호랑이 존재를 알아차리고 밖으로 탈출한다	9. 남매가 나무 위로 숨는다	10. 호랑이는 올라오는 방법을 묻는다	11. 아이들이 응답하거나 호랑이를 속인다	12. 남매가 하늘에 반다	13. 줄이 내려와 남매가 하늘로 올라간다	14. 남매는 해와 달이 된다	15. 호랑이도 하늘에 빌어 줄을 타고 올라간다	16. 줄이 끊어져 호랑이가 수수대밭에 추락한다	17. 호랑이의 피가 묻어 수수대가 붉은 것이다	변주 양상
1	해와 달이 된 두 아이	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수밭/수수대 추락
2	해와 달이 된 두 아이	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수밭/수수대 추락
3	해와 달이 된 두 아이	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수밭/수수대 추락
4	해와 달이 된 두 아이	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수밭/수수대 추락
5	해와 달	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수밭/수수대 추락
6	수수대밭 빨간 줄	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수밭/수수대 추락
7	해와 달이 된 두 아이	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수대 추락
8	해와 달이 된 두 아이	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	수수대 추락

[그림 4] 사건 항목별 구조화 표

위 [그림 4]의 표를 살펴보면 각편별로 사건의 출현 여부와 변주 지점을 정리함으로써 동일 유형의 이야기들을 비교 가능한 방식으로 배열하고 있음을 알 수 있다. 각편에 따른 변주 양상을 구체적으로 확인하기 위해 [그림 4]에서 변주 요약 부분을 확대하여 제시하면 다음 [그림 5]와 같다.

번호	제목	변주 요약
3	해와 달이 된 오누이	3. 호랑이가 먹을거리나 신체 일부를 요구한다=떡, 옷; 6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다 =목소리, 손-애기/젖; 8. 남매는 호랑이 존재를 알아차리고 밖으로 탈출한다=뒋으로 탈출; 11. 아이들이 응대하거나 호랑이를 속인다=기름→도끼/자귀; 14. 남매는 해와 달이 된다=교체 요청 불가, 오빠 해; 16. 줄이 끊어져 호랑이가 수숫대밭에 추락한다=수수밭/수숫대 추락
4	해와 달이 된 오누이	3. 호랑이가 먹을거리나 신체 일부를 요구한다=메밀묵/묵, 옷, 팔/다리, 함지/함박; 4. 어머니가 호랑이에게 잡아 먹힌다=팔다리 절단 뒤 잡아먹음; 6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다=목소리, 손-애기/젖; 9. 남매가 나무 위로 숨는다=넋이나무; 11. 아이들이 응대하거나 호랑이를 속인다=기름→도끼/자귀; 13. 줄이 내려와 남매가 하늘로 올라간다=줄, 방석; 14. 남매는 해와 달이 된다=해, 달 교체, 오빠 달; 16. 줄이 끊어져 호랑이가 수숫대밭에 추락한다=수수밭/수숫대 추락
5	해와 달	6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다=목소리, 손; 8. 남매는 호랑이 존재를 알아차리고 밖으로 탈출한다=배편 핑계로 탈출; 11. 아이들이 응대하거나 호랑이를 속인다=기름→도끼/자귀; 16. 줄이 끊어져 호랑이가 수숫대밭에 추락한다=수수밭/수숫대 추락
6	수숫대가 빨간 유래	3. 호랑이가 먹을거리나 신체 일부를 요구한다=떡, 팔/다리; 6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다=손; 9. 남매가 나무 위로 숨는다=오동나무; 11. 아이들이 응대하거나 호랑이를 속인다=기름→도끼/자귀; 14. 남매는 해와 달이 된다=오빠 달, 동생/누이 해; 16. 줄이 끊어져 호랑이가 수숫대밭에 추락한다=수수밭/수숫대 추락
7	해와 달이 된 오누이	6. 아이들은 어머니임을 확인할 수 있는 표지를 요구한다=목소리, 손; 9. 남매가 나무 위로 숨는다=계수/계수나무; 11. 아이들이 응대하거나 호랑이를 속인다=기름→도끼/자귀; 14. 남매는 해와 달이 된다=오빠 달; 16. 줄이 끊어져 호랑이가 수숫대밭에 추락한다=대밭/수숫대 추락

[그림 5] 각편별 변주 지점

[그림 5]는 [그림 4]의 전체 구조화표 가운데 3~7번 각편의 제목과 변주 요약 부분만을 따로 제시한 것으로, 각편별 변주가 어느 사건항에서 발생하는지를 보다 분명하게 보여준다. 예컨대 각편 <4>³⁰⁾의 경우 3항이 다른 각편에 비해 구체적으로 제시되어 있다. 즉 이 각편에서는 호랑이가 어머니에게 ‘메밀묵’, ‘옷’, ‘팔’, ‘다리’, ‘함박’을 요구한 것으로 나타나는데, 이는 이 각편을 구연한 화자가 다른 화자들에 비해 호랑이가 어머니를 잡아먹기까지의 과정을 중점적으로 구연했음을 짐작할 수 있는 부분이다. 또한 13항에서 확인되듯이 남매가 하늘로 올라가는 부분에서 ‘줄’뿐 아니라 ‘방석’도 등장했음을 알 수 있다. 이처럼 사건 단위 기반 구조화는 각편별 변주 지점을 제시해줌으로써 비교를 위한 분석 좌표로 기능한다.

이러한 구조화의 효용은 일단 자료 식별 차원에서 확인된다. 즉 동일

30) 이후 서술부터는 각편의 경우 사건항의 번호와 구분하기 위해 홑화살표호(< >)를 사용하여 표기하기로 한다.

유형의 서로 다른 제목의 자료일지라도 사건 단위를 기준으로 배열하면 1~17항의 사건을 공유하는 자료로 다시 식별된다. 이는 사건 단위 구조화가 표제어와 같은 자료만을 찾아내는 데에 그치지 않고 제목이 다르더라도 동일한 사건 배열을 공유하는 자료를 판별할 수 있게 한다는 점에서 의미가 있다.

또한 이러한 구조화는 변주 지점을 비교하는 데에도 효과적이다. [그림 5]에서 확인되듯 각편별 변주 지점은 모든 사건항이 아닌 특정 사건항에 집중되어 나타난다. 예컨대 호랑이가 어머니에게 요구하는 대상은 떡, 밥, 메밀묵, 옷, 팔, 다리 등으로 다양하게 나타나고, 남매가 어머니임을 확인하는 방식 역시 목소리, 손, 털, 발 모양 등으로 변주된다. 이처럼 사건 단위 기반 구조화는 각편에서 변주가 발생하는 사건항을 더욱 분명히 드러내주는 것이다.

다만 이러한 AI의 산출 결과 역시 검증이 필요하다. AI가 제공한 구조화 결과 자료와 원문 각편을 대조해본 결과 해당 사건이 등장하는데도 불구하고 사건항이 공란으로 처리된 사례도 포착되었다(그림 4).³¹⁾ 이는 AI가 제공한 자료는 반드시 원문과 대조하여 항목을 조정하는 과정이 필요함을 보여준다. 그런데 여기서 특기할 만한 점은 AI가 구조화한 결과는 없는 내용을 기입하기보다는³²⁾ 있는 내용을 누락한 경우가 더 빈번하였는데 이는 특히 표에서 공란으로 되어 있는 곳을 집중적으로 검토해야 함을 드러낸다. 즉 구조화표에서 빈칸이 실제로 사건이 생략된 것인지 혹은 AI가

31) 예컨대 각편 <5>에서 호랑이가 먹을 것을 요구하는 사건항³이 나타나지 않는다고 제시되어 있지만 원문을 보면 밥을 달라는 내용을 찾아볼 수 있으며, 호랑이의 기원으로 동아줄이 내려온 후 타고 올라가는 사건항¹⁵도 마찬가지로 나타나는 것을 확인할 수 있다.

32) 이는 다른 부분에서는 거의 찾아볼 수 없고 남매가 어머니의 정체확인을 요청하는 사건항⁶에서 변주 지점으로 ‘젓’으로 제시한 부분에서 나타난다. ‘애기(갓난아이)’가 등장하는 경우 ‘젓’도 함께 등장함을 확인할 수 있었다.

누락시킨 것인지를 확인해야 하는 중점적 검토 지점인 것이다. AI를 활용한 구조화는 이러한 비교 연구의 예비 자료로 이용될 수 있을 뿐 아니라 다양한 활용 가능성이 있다.

먼저 자료의 재배열을 가능하게 할 수 있다. 즉 사건 단위 기반 구조화가 이루어지면 동일 자료를 원하는 범주로 다시 묶을 수 있는 것이다. 이를테면 ‘정체 확인 방식이 손으로 나타나는 각편을 찾아줘’, ‘해와 달의 배경이 교체되는 각편을 찾아줘’, ‘수숫대 유래가 포함된 각편을 찾아줘’와 같은 질문을 AI가 안정적으로 처리하여 재배열할 수 있게 된다. 이는 사건 단위의 구조화가 AI의 자연어 처리 결과를 비교 가능한 데이터 처리 결과로 전환하는 기반이 된다는 것이다.

이처럼 사건 단위 구조화는 <해와 달이 된 오누이> 유형의 각편들을 동일한 사건 배열 아래 놓고 비교를 가능하게 해준다. 특히 표제가 다른 자료들도 공통 사건 단위를 기준으로 다시 식별할 수 있고, 각편별 변주 역시 특정 사건 단위의 실현 방식의 차이로 파악하게 해준다는 점에서 의미가 있다. 다만 AI가 제공한 결과는 원문과의 대조를 통해 검토하고 보정하는 과정이 반드시 수반되어야 한다. 이러한 맥락에서 본고의 작업은 AI의 예비 처리 결과를 인간 연구자가 검토·보정하는 Human-in-the-Loop 기반 구조화 방식이라고 할 수 있다. 이를 통해 비정형의 구술 채록 자산인 <한국구비문학대계> 자료는 비로소 일관된 기준에 따라 정렬되어 비교 가능한 데이터로서의 실체성을 획득할 수 있다.

5. 결론

본고는 <한국구비문학대계> 디지털 아카이브에 수록된 구비설화 자료를 생성형 AI 활용 환경에서 보다 안정적으로 식별·비교·재배열하기 위한

구조화 방안을 검토하였다. 이러한 작업은 생성형 AI가 구비설화를 비롯한 비정형 고전문학 자료를 연구 대상으로 다룰 때, 단순한 검색이나 요약은 넘어 자료의 구조를 파악하고 비교 가능한 형식으로 전환하는 데 어떻게 활용될 수 있는지를 구체적으로 보여준다는 점에서 의미가 있다. 본고는 하나의 설화 유형만을 대상으로 사건 단위 설정 기준을 마련하고, 이를 각 편에 적용하여 각편별 비교를 가능하게 하는 구조화 단계까지만 살펴보았지만, 이를 바탕으로 구조화표를 더욱 정교한 연구 데이터로 확장하는 방향으로 나아갈 수 있다.³³⁾

이들테면 각편 제목, 제보자 성별과 연령, 채록 지역, 수록 위치, 원문 URL 등과 같은 메타데이터를 사건 단위 구조화표와 연결할 수 있다. 이러한 메타데이터 연계 구조가 확보되면 화자의 성별, 지역에 따른 전승 양상의 차이 등을 다각도로 분석할 수 있다. 또한 본고에서는 각 사건항에 따른 변주 양상을 변주 요약으로 제시하였지만 이를 더 세분화하여 구조화할 수도 있다. 나아가 이러한 구조화는 다른 설화 유형에도 적용할 수 있다. 이와 같은 작업이 축적된다면 <한국구비문학대계>는 단순한 원문 아카이브를 넘어 유형별 사건 구조와 변주 양상을 비교할 수 있는 지식 기반으로 확장될 수 있을 것이다.

요컨대 본고에서 살핀 ‘사건 단위 스키마’와 인간 연구자-AI 협업 기반 구조화 파이프라인은 대규모 구비설화 아카이브 자료를 반자동으로 구조화하고 검토하기 위한 기초 데이터 형식으로 기능할 수 있다. 즉 이 구조화 데이터를 참조 예시로 활용하여 프롬프트를 고도화함으로써 다른 설화 유형의 사건 단위와 표제어를 체계적으로 분류하는 전문가 검증 기반의 반자동 파이프라인 구축이 가능해질 수 있다. 또한 이를 기반으로 세계 각국의 구비문학 데이터를 다차원적으로 매핑한다면 국제 비교문학 연구 역시 한

33) 이는 후속 논고에서 다루고자 한다.

층 입체적이고 정밀해질 수 있을 것이다. 결과적으로 이러한 지식 기반으로의 확장은 인문학이 기술을 도구적으로 수용하는 일방향적 활용을 넘어 구비문학 자료가 거대언어모델을 한국적 맥락에 맞게 미세조정하거나 정렬하는 학습 데이터로 기능함으로써 모델의 성능을 고도화하고 문화적 편향을 교정하는 역방향적 기여의 가능성을 시사한다는 점에서도 의미가 있다. 본고가 제안한 인간 연구자-AI 협업 기반의 서사적 스키마가 지닌 학술·기술적 확장성에 대해서는 향후 후속 논고를 통해 규명하고자 한다.

참고문헌

- 한국구비문학대계 디지털 아카이브, <https://kdp.aks.ac.kr/inde/gubi>.
- OpenAI, ChatGPT(GPT-5.4 Thinking, 활용일자 2024.4.15.), 대형언어모델(LLM), <https://chatgpt.com/>, 2026.
- 김성우, 『인공지능은 나의 읽기-쓰기를 어떻게 바꿀까』[전자책], 유투, 2024.
- 김정은, 「『구비문학연구』 30년, 민담분석의 성과와 과제」, 『구비문학연구』 70, 한국구비문학회, 2023, 91~139쪽.
- 김정희, 「담춘 선생의 서사 구조와 의미 분석 방법을 활용한 검색 시스템의 설계-건국신화와 구비설화 연구를 중심으로」, 『구비문학연구』 63, 한국구비문학회 2021, 31~68쪽.
- 문학치료 서사사전 연구모임, 『문학치료 서사사전2』[전자책], 휴머니스트, 2002.
- 박태웅, 『박태웅의 AI 강의 2025』[전자책], 한빛비즈, 2024.
- 박현숙, 「설화 아카이브의 현황 분석과 활용 방안」, 『구비문학연구』 55, 한국구비문학회, 2019, 209~252쪽.
- 블라디미르 프로프, 『민담형태론』, 유영대 옮김, 새문사, 2009, 1~200쪽.
- 서대석, 『이야기의 의미와 해석』, 세창출판사, 2020, 1~395쪽.
- 신동훈, 「서사적 화두를 축으로 한 화소·구조 통합형 설화분석 방법 연구」, 『구비문학연구』 46, 한국구비문학회, 2017, 35~84쪽.
- 심우장, 「설화에 나타난 서사 확장의 두 패턴」, 『구비문학연구』 65, 한국구비문학회 2022, 113~152쪽.
- 이승은, 「고전서사 플랫폼 구축 시론-『한국구비문학대계』 디지털 플랫폼 프로토타입 제작과 과제-」, 『고전과 해석』 48, 고전문학한문학회, 2026, 35~68쪽.
- 한유진, 「한국 고전문학의 AI 번역, 그 가능성과 한계-〈아랑〉 설화 번역의 문화적 굴절을 중심으로-」, 『온지논총』 84, 한국온지학회, 2025, 249~281쪽.
- Dundes, Alan, "From Eitic to Emic Units in the Structural Study of Folktales", *Journal of American Folklore*, vol. 75, No. 296, 1962, pp.95~105.
- Dundes, Alan, "Structural Typology in North American Indian Folktales", *The Study of Folklore*, ed. Alan Dundes, Englewood Cliffs, N.J. : Prentice-Hall, 1965, pp.206~215.

OpenAI(2026.3.18.), “Model Release Notes”, https://help.openai.com/en/articles/9624314-model-release-notes?utm_source=chatgpt.com, 접속일 2026.5.17.

ABSTRACT

A Study on the Structuralization
of Korean Oral Folktale Materials Using Generative AI
-Focusing on Materials from The Great Collection of Korean Oral Literature-

Han, Yu-jin

This study examines a method for structuralizing Korean oral folktale materials contained in the digital archive of The Great Collection of Korean Oral Literature so that they can be more stably identified, compared, and rearranged in an environment where generative AI is used. To this end, this study focuses on the tale type *The Brother and Sister Who Became the Sun and the Moon*. First, relevant materials were searched using AI, and the researcher then reviewed the results and confirmed forty-five texts as the object of analysis. Based on these forty-five texts, the preliminary structuralization results suggested by AI were examined. Items involving interpretive naming or inconsistent levels of classification were adjusted, and criteria for establishing event units were developed around observable actions and situations.

On this basis, the narrative was organized into seventeen event units: the mother's departure, the encounter with the tiger, the tiger's demand, the mother's death, the tiger's attempted intrusion, the children's verification of whether the visitor is their mother, the tiger's intrusion, the children's escape, the children's hiding, the tiger's pursuit, the children's response or deception, the children's prayer, the children's ascent, the transformation into the sun and the moon, the tiger's attempted ascent, the tiger's fall, and the origin of the red sorghum stalks. These criteria were applied to all forty-five texts. Events that appeared repeatedly were marked in the same way, while variant points in each text were recorded separately. The results show that event-unit-based structuralization enables materials with different titles

to be identified again on the basis of their shared event sequences. It also makes it possible to compare variants as differences in the way specific event units are realized and to rearrange the materials according to various analytical categories. However, because AI-generated structuralization may include omissions or other errors, comparison with the original texts and correction by the researcher are essential.

This study is significant in that it presents a methodological possibility for transforming Korean oral folktale materials into research data that can be compared and accumulated through the use of generative AI. As a data-processing method that combines AI-based preliminary processing with human review and reconfiguration, this approach suggests that large-scale oral folktale collections such as The Great Collection of Korean Oral Literature can be expanded beyond a simple archive of original texts into a knowledge base for comparing event structures and patterns of variation by tale type. Furthermore, the ‘event unit schema’ and human-AI collaboration pipeline examined in this study demonstrate potential for systematic expansion into other voluminous tale types and international comparative literature.

Key Words

The Great Collection of Korean Oral Literature, Korean oral folktales, generative AI, structuralization, *The Brother and Sister Who Became the Sun and the Moon*, event unit schema, Human-in-the-Loop, Digital Humanities

논문투고일: 2026.04.23.

심사완료일: 2026.05.13.

게재확정일: 2026.05.18.