

컴퓨터 문헌 분석 기법을 활용한 <적벽가> 이본의 계통 분류 연구

A Study on the classification of Jeokbyeok-ga's version by the Computer analysis technique of bibliographies

이진오*, 김동건**
 금오공과대학교*, 경희대학교**

Lee Jin-O*, Kim Dong-Keon**
 Kumoh National Institute of Technology*,
 KyungHee University**

요약

본 연구는 컴퓨터 문헌 분석 방법을 활용하여 <적벽가> 이본의 계통을 분류하고 기존 이본 연구를 검토한 것이다. <적벽가> 이본의 공통 서사단위는 5개의 계층으로 파악되며, 146개의 개별단락을 산출해낼 수 있다. 이를 통해 군집 분석 방법과 계통의 분기 분석 방법을 적용할 수 있으며, 작품의 계통과 이본간의 거리를 시각적으로 파악할 수 있다.

I. 서론

<적벽가>는 『삼국지연의』를 기반으로 한 작품군으로 창본과 소설본 형태로 다양한 이본을 남겼다. <적벽가>의 성립과 이본 계통을 고찰한 연구는 여러 연구자에 의해 진행되었으나 직관에 의존한 면이 있다.

본 연구에서는 내용 단락 유사도에 기반한 컴퓨터 문헌 분석 방법론을 <적벽가>에 적용해보고 기존 연구의 계통 분류 결과를 검토하고자 한다.

II. 자료의 내용 분석과 단락별 유형 분류

1. 자료의 유형 인코딩

본 연구는 『적벽가 전집』 1~7권에 수록된 작품들을 기초 자료로 삼는다. 『적벽가 전집』에는 총 54종의 이본이 실려 있는데, <적벽가>의 계통과 특징을 뚜렷이 보여줄 수 있는 이본 46종을 연구대상으로 선별하였다.

컴퓨터 문헌 분석 방법론의 기반 작업을 마련하기 위해서 <적벽가>의 서사 단락과 단락별 유형이 반영된 작품 전체 이본의 데이터화 작업을 수행하였다. <적벽가>의 이본에는 형태에 따라 C, P, S, E의 식별 ID를 부여하였고, 각 이본의 내용 단락의 유형을 분류하는 작업을 진행하였다. 그 다음으로 분류한 단락을 전산화된 원시 코퍼스로 구성하였고, 유형별로 인코딩하였다.

2. 내용 단락과 유형 분류

<적벽가>의 각 이본에 따라 개별 단락의 미세한 차이까지 측정하기 위해 공통 서사 단락을 '대단락(L1)', '중단

락(L2)', '소단락(L3)', '소소단락(L4)', '소소소단락(L5)'의 계층 구조로 분석하였다. 이는 내용 단락의 상호 대비와 관계를 살피기 위해 상위의 서사 단락을 인물과 사건의 분화에 따라 세분화한 것이다.

다음으로 <적벽가> 전체 이본의 각 내용 단락에 인코딩한 기본 자료를 유형별로 분류하였다. 이후 각 작품의 유형 표지를 추출하였고, 변별되는 내용 단락을 기호로 나타낸 작업 테이블을 구축하였다.

SID	ID										
	1	2	3	4	5	42	43	44	45	46	
EUJN02	SDGL02	SDGL03	SDGL04	SKKS02	...	CSS5	CHSH	CJK	CKYS	PYIB	
S0001	a	a	a	a	a	...	b1	d	b5	b2	x
S0002	x	x	x	x	x	...	x	x	x	x	x
S0003	a1	a1	a1	a1	a1	...	b1	b1	b2	b2	a1
S0004	x	x	x	x	x	...	x	x	b1	b1	x
S0005	x	x	x	x	x	...	b2	b2	b4	b4	x
S0006	a	a	a	a	a	...	x	x	x	x	x
S0007	x	x	x	x	x	...	b2	b2	b1	b1	x
S0008	x	x	x	x	x	...	b1	b1	b1	b1	x
S0009	x	x	x	x	x	...	b2	b2	b1	b1	x
S0010	a1	a2	a1	a1	a1	...	b1	b6	b7	b1	b2
S0011	a1	a1	a1	a1	a1	...	a2	a2	b	a2	x
S0012	a1	a1	a1	a1	a1	...	d	a3	c	c	x
S0013	x	x	x	x	x	...	x	x	c	c	x
S0014	x	x	x	x	x	...	x	o	x	x	x
S0015	x	x	x	x	x	...	x	b1	x	x	x
S0135	a1	a2	a1	a1	낙	...	b1	x	x	b3	x
S0136	a1	b	a1	a1	낙	...	c1	c3	d2	e	a3
S0137	a	x	a	a	낙	...	b2	x	x	b2	x
S0138	c	x	a	a	낙	...	d1	b1	b1	b2	g
S0139	a	a	a	a	낙	...	c1	c1	c1	c4	e2
S0140	a	b	b	b	낙	...	x	x	x	x	x
S0141	o	o	o	o	낙	...	x	x	x	x	x
S0142	o	o	o	o	낙	...	x	x	x	x	x
S0143	o	x	o	o	낙	...	x	x	x	x	x
S0144	o	o	o	o	낙	...	x	x	x	x	x
S0145	o	x	o	o	낙	...	x	x	x	x	x
S0146	x	x	x	x	낙	...	a2	a2	a2	c	x
S0147	o	x	o	o	낙	...	x	x	x	x	o
S0148	x	x	x	x	낙	...	a2	a2	a2	c	x
S0149	o	x	o	o	낙	...	x	x	x	x	o

▶▶ 그림 1. 이본별 내용 단락 유형 테이블

그림 1은 <적벽가>의 이본별 내용 단락을 유형화 하여 정리한 테이블로 각 내용 단락마다 고유의 일련번호가 부여되었음을 보여준다. 작품의 내용 단락 유형은 알파

