

‘춘향가’ 서두 단락의 어휘 사용 유사도를 이용한 판본 계통 분류 연구

2009, 2011년 연구 자료

연구 목적, 방법

▶ 목적

- ▶ ‘춘향가’ 이본의 계통 분류

▶ 방법론적 기초

- ▶ Computational Approaches to Textual Criticism

- ▶ 텍스트와 텍스트의 유사/차이 측정 방법

- ▶ 최소편집거리/레벤슈타인거리

- 텍스트의 표기, 문법형태의 차이 등의 비교
 - ‘전승과 변모’의 양상을 ‘거리’로 측정
 - 최운호, 김동건 (2009): ‘수궁가’의 전승과 변모

- ▶ 어휘 사용의 유사도

- ‘lower criticism (변이 양상: 철자 변이 등)’을 어휘 사용으로 환원
 - ‘higher criticism (저자 판별, 작품의 계보 등)’에 대한 해석에 사용
 - 거리 함수: Jaccard Similarity



자료 소개와 표기법 정규화 (1/2)

□ 『춘향전 전집』 1~17권 수록 춘향전 이본 작품 전체의 서두 단락

- 춘향전 이본 자료 총 108종 수록
 - 판각본 10종, 필사본 70종, 활자본 11종, 창본 18종
- 총 83종 분석 대상
 - 서두 탈락 이본이나 교합본 제외
 - 총 83종 선별
 - 판각본 10종
 - 필사본 52종 (경판 5종, 안성판 1종, 완판 4종)
 - 창본 14종
 - 활자본 7종

□ 자료의 변환과 정규화

- 전자법(transliteration)에 의한 현대어 변환
- 형태 주석 말뭉치 구축 (HAM[2] & 사후 교정)



자료 소개와 표기법 정규화 (2/2)

ID: SKDU01, 김동욱 소장 낙장70장본 <춘향전>

숙종대왕 즉위 초의 국태민안하고 시화연풍이라 조정에는 충신이요
여염의난 효조 열여료다 百姓而 壑壤歌乙 일삼을 제

현대어 변환 자료

숙종대왕 즉위 초에 국태민안하고 시화연풍이라 조정에는 충신이요
여염에는 효자 열녀로다 백성이 격양가를 일삼을 제

형태 주석 말뭉치

숙종대왕	숙종대왕/N
즉위	즉위/N
초에	초/N + 예/j
국태민안하고	국태민안/N + 하/t + 고/e
시화연풍이라	시화연풍/N + 이/c + 으라/e
조정에는	조정/N + 예는/j
충신이요	충신/N + 이/c + 요/e
여염에는	여염/N + 예는/j
효자	효자/N
열녀로다	열녀/N + 이/c + 로다/e
백성이	백성/N + 이/j
격양가를	격양가/N + 을/j
일삼을	일삼/V + 을/e
제	제/N

► Jaccard Sim.

$$\text{dist}_{\text{Jaccard}}(\vec{v}, \vec{w}) = 1 - \text{sim}_{\text{Jaccard}}(\vec{v}, \vec{w}) \quad (2)$$

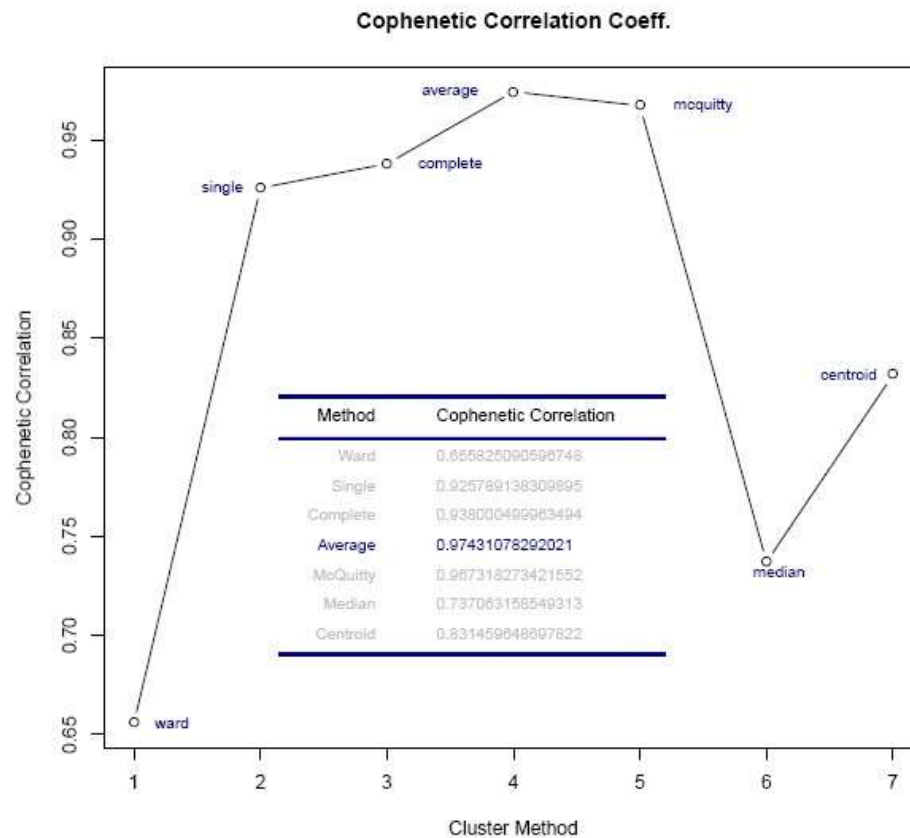
The graph shows the cumulative distribution of PBKS values. The x-axis represents the PBKS value, ranging from 0.000000 to 1.000000. The y-axis represents the cumulative probability, ranging from 0.0 to 1.0. The curve starts at (0.000000, 0.0) and rises sharply, reaching a cumulative probability of 1.0 at a PBKS value of approximately 0.980000. The curve is smooth and S-shaped, indicating a non-uniform distribution.

▶ 총 83종 이본에 대한 비교

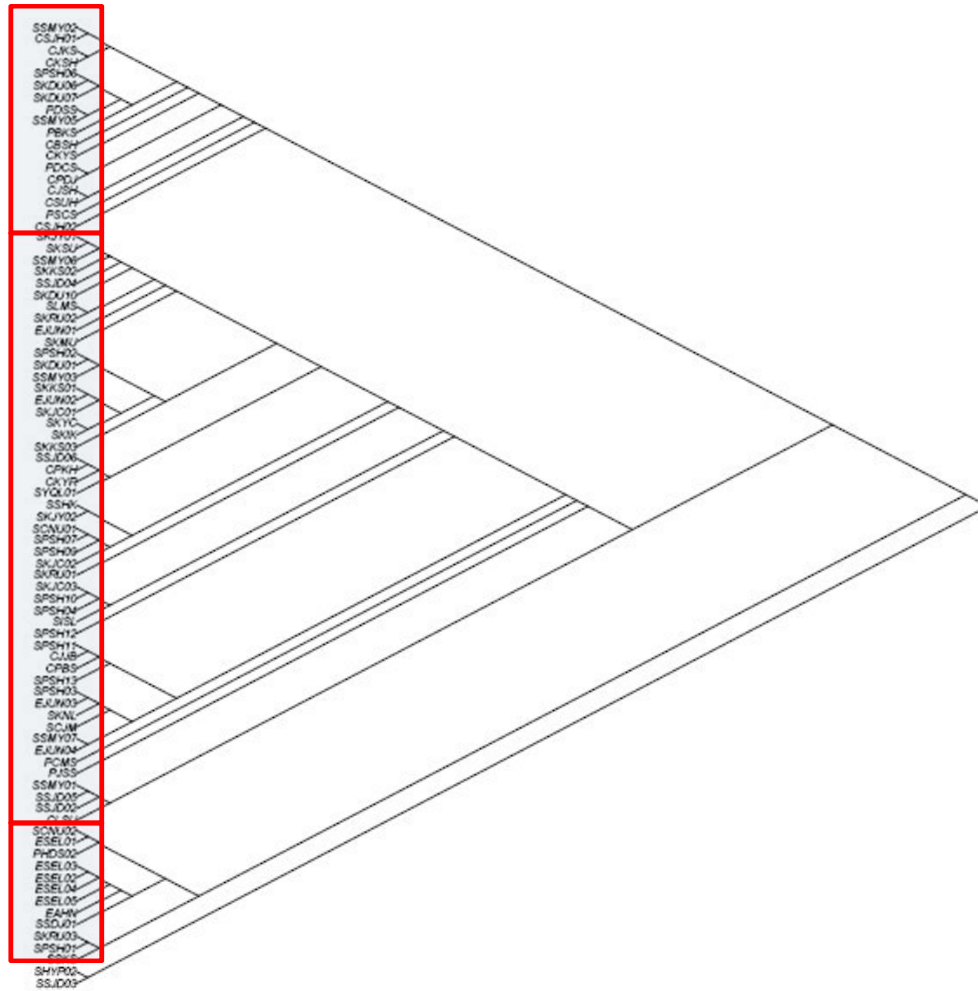
[illegible]

Cophenetic 상관계수

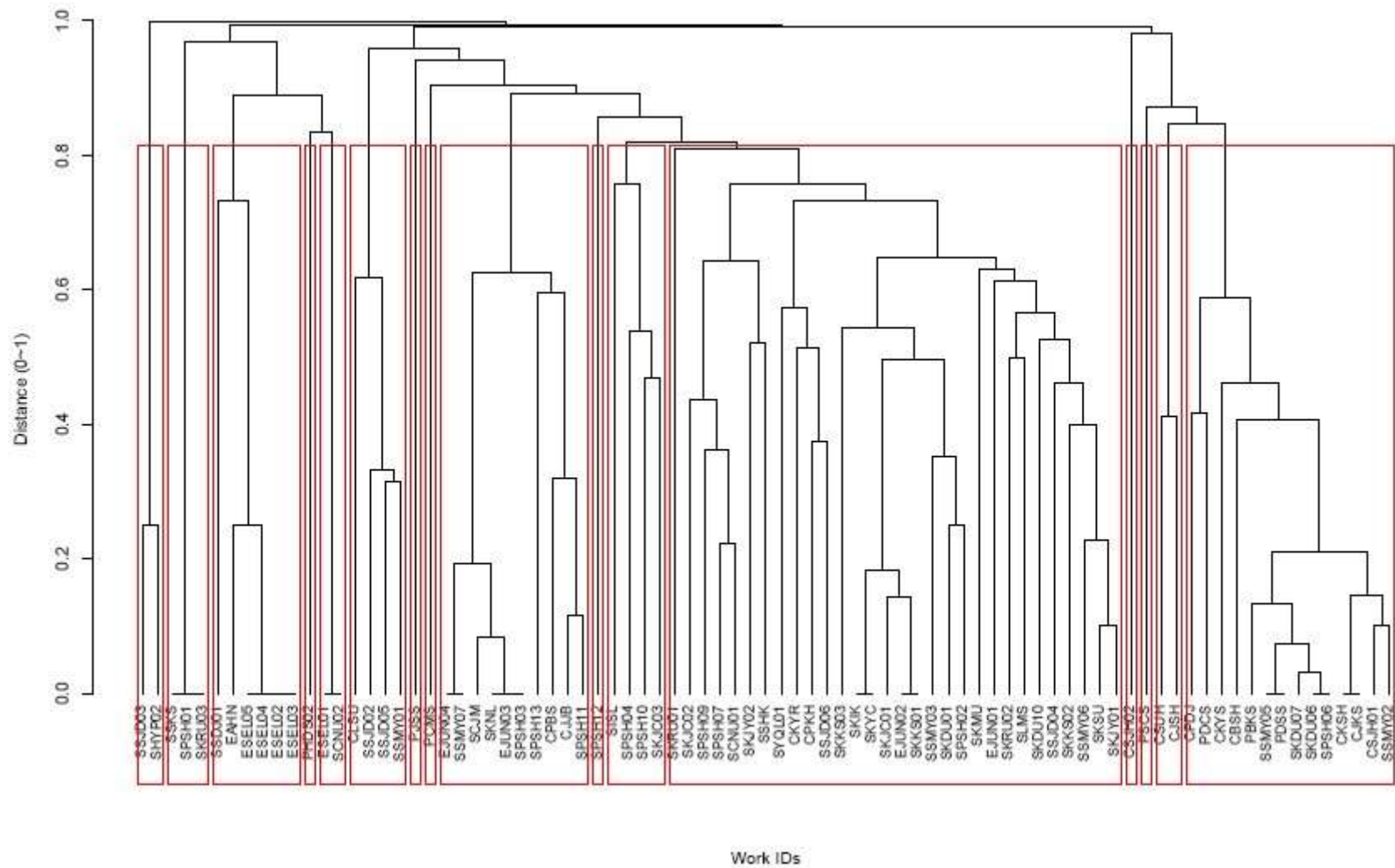
- ▶ 계층적 군집분석 방법의 선택
 - ▶ 거리행렬과 군집분석 결과의 적합도



결과 (1 / 2), left-clade



결과 (2/2), hclust (16 cut)



결론

▶ 연구의 의의

- ▶ 이본에 사용된 어휘 형태 (사용) 분포의 유사성에 기초해서 이본들 사이의 (내용) 차이를 척도화 하는 모델 적용
- ▶ 기존 이본 연구
 - ▶ 연구자의 (주관에 따라) 분류하던 방식을 개선, 객관성 부여
 - ▶ 이본 사이의 분류뿐만 아니라, 그 차이도 척도화

