

다중 네트워크 분석과 토픽 모델링을 이용한 임진왜란 시기 사료에 관한 연구*

A Study on the Imjin War's Historical Materials with Multi-layer Network Analysis and Topic Modeling

조 현 철 (HyunChul Cho)**

송 민 (Min Song)***

초 록

융합 과학 연구가 활성화되며 인문학에서도 디지털 인문학(Digital Humanities) 연구가 장려되고 있다. 이에 본 연구는 역사 데이터에 텍스트마이닝과 개체계량학 연구 방법을 적용한 시론(試論) 연구를 제안하고자 하였다. 선조실록(宣祖實錄)·선조수정실록(宣祖修正實錄), 난중잡록(亂中雜錄), 징비록(懲毖錄)을 활용하였으며, 사료(史料)에서 주제 변화와 공통 개체를 탐색하기 위해서 네트워크 분석과 DMR 토픽모델을 사용하였다. 분석 결과를 통해서 텍스트 데이터에 대한 계량 분석의 활용 가능성 확인, 특정 주제의 시기적 변화, 인물 개체 간 미발견 관계를 제시함으로써 연구의 확장 가능성을 제안할 수 있었다.

ABSTRACT

Convergence science research is activated, and digital humanities research is also encouraged in humanities. Therefore, this study attempted to propose a experimental study that applies Text mining and Entitymetrics methods to historical materials. Annals of King Seonjo, revised Annals of King Seonjo, Miscellaneous Record of the War and Writings on Imjin War were used, also network analysis and DMR topic models were used to explore topic changes and common entities in historical sources. Through the results, it was possible to propose the availability of quantitative analysis for text data, presenting a timing change of a specific topic, and an undiscovered relationship between person entities.

키워드: 디지털 인문학, 개체계량학, 조선왕조실록, 난중잡록, 징비록, DMR 토픽 모델링, 다중 네트워크
Digital Humanities, Entitymetrics, Annals of the Joseon Dynasty, Miscellaneous Record of the War, Writings on Imjin War, DMR topic modeling, Multi layer network

* 이 논문은 석사학위논문(2021년 12월)의 축약본임.

이 논문은 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A3A2075114).

** 연세대학교 일반대학원 문헌정보학과 석사과정(chhyunchul94@gmail.com) (제1저자)

*** 연세대학교 문헌정보학과 교수(min.song@yonsei.ac.kr) (교신저자)

논문접수일자 : 2022년 2월 21일 논문심사일자 : 2022년 2월 24일 게재확정일자 : 2022년 3월 3일
한국비블리아학회지, 33(1): 167-198, 2022. <http://dx.doi.org/10.14699/kbiblia.2022.33.1.167>

© Copyright 2022 Korean Biblia Society for Library and Information Science

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

디지털 인문학은 다양한 학문의 연구 방법을 융합하여 기존 연구에서 발견하지 못한 연구 주제와 관계를 도출할 수 있다는 장점이 있다. 디지털 인문학은 인문학 분야와 정보학, 사회학, 공학 등 다양한 학문과 접목할 수 있다. 또한, 디지털 인문학 연구의 기초 단계인 데이터 전산화는 연구자와 대중의 자료 접근성을 향상한다. 디지털 인문학은 인문학의 새로운 연구 방법으로써 인문학의 연구 대상을 유지하며 다양한 분석 방법을 응용하는 보조적인 연구다(김현, 임영상, 김바로, 2016).

김바로(2014)는 디지털 인문학이 가진 융합 학문의 성격을 '인문학의 위기'의 해결책으로 보았다. 기존 인문학이 고도화되며 이해의 난해와 함께 연구성과가 사회에 환원되기 어려워진 문제를 해결하리라 보았다. Burdick et al.(2012)은 디지털 인문학이 전산화된 자료 이용과 고정된 데이터의 형태를 변형하는 특징을 갖고 있다고 하였다. 이처럼 디지털 인문학은 실물로 고정된 데이터를 전산화 작업을 거침으로써 연구자와 대중들의 인문학 자료에 관한 관심을 높임으로써 기존 인문학 연구를 보완하는 효용성을 지니고 있다.

디지털 인문학 연구는 조지형(2001)에 따르면 2000년대의 정보화 사회 도래와 함께 인문학이 맞이한 위기를 해결할 대안으로써 화두가 되었다. 국내 디지털 인문학 연구의 시작은 1967년에 시작된 에드워드 와그너(Edward W. Wagner, 1967-2001)와 송준호(宋俊浩, 1922-2003)의 '문과방목(文科榜目) 프로젝트'로 볼 수 있다. 국사편찬위원회에서는 1993년에 조선왕조실록

(朝鮮王朝實錄)의 전산화 사업으로 국역본과 한문 원본을 CD-ROM으로 제작하여 연구자와 대중의 사료에 대한 접근성을 높였다. 실록을 시작으로 해서 인문학 데이터에 대한 전산화 작업은 국사편찬위원회, 한국학중앙연구원, 한국고전번역원을 중심으로 진행되었다. 대표적인 성과로는 연표, 독립운동사 자료, 민주화운동 자료, 비변사등록, 고려사, 삼국사기의 데이터 베이스화가 있다. 주성지(2018)는 이러한 전산화 사업이 2001년부터 2010년까지의 한국사 연구 성과와 역사 콘텐츠 제작 현황에 큰 영향을 주었다고 분석하였다.

인문학 데이터에 대한 전산화 작업은 다양한 종류에 대해서 이루어졌으나, 현행 디지털 인문학 연구는 국내외의 연구 동향 분석이 주를 이루고 있다. 대표적으로 역사학 분야는 '역사학보'를 디지털 인문학 특집호로 구성하여 세계 디지털 인문학 연구 동향을 다루었다. '각국의 디지털 역사학'을 소주제로 미국, 영국, 대만, 프랑스의 디지털 인문학 연구에 대한 4편의 논고가 게재되었다. 해당 특집호에서는 국가별 디지털 인문학 연구 변천, 현행 과제와 연구를 상술하였다. 또한, 이재연(2018)은 한국문학을 중심으로 국내 연구 현황과 함께 디지털 인문학 연구사의 변천을 정리하였다(권윤경, 2018; 문수현, 2018; 박은재, 2018; 송인재, 2018; 이재연, 2018; 이주영, 2018).

정보학 분야에서는 정유경(2020)과 정은경(2021)은 국내외 디지털 인문학 학술지와 학술대회를 중심으로 디지털 인문학 연구 동향을 분석하였다. 정유경(2020)은 국내외 디지털 인문학 연구 동향 비교를 위해서 Web of Science와 RISS의 문헌을 수집하여 구조적 토픽모델

을 사용한 주제 분석을 수행했다. 정은경(2021)은 Digital Humanities 학술대회 논문을 대상으로 디지털 인문학 연구 경향을 분석하였다.

이처럼 국내 디지털 인문학은 동향 파악을 중심으로 연구가 이루어지고 있으나 조원희(2017)와 김바로(2021)의 활용 연구 사례도 있다. 조원희(2017)는 원인종(元仁宗, 1285-1320)이 재위하던 시기의 조정 신료 네트워크를 분석하여 기존 연구에서는 주목하지 않았던 하산(阿散, 1255-1304)에 대한 새로운 관점을 제시하였다. 김바로(2021)는 딥러닝 언어 모델인 BERT (Bidirectional Encoder Representations from Transformers)를 사용하여 개체명 인식(Named Entity Recognition)을 활용하여 사전 구축 연구를 하였다. 해당 연구는 역사학 데이터에 딥러닝 모델을 사용함으로써 향후 지향점을 ‘기계 가독형 데이터’를 활용한 개체 간 연결 관계 구축으로 보며 번역과 문헌 분류를 중심으로 사용되던 디지털 인문학에 다양한 분석 도구로써 딥러닝 활용 가능성을 보였다.

본 연구의 목적은 연구 동향 분석이 중심을 차지하고 있는 경향을 넘어서 계량 분석 방법을 기반으로 한 구체적인 디지털 인문학 접근 방향 제안에 있다. 융합을 통하여 인문학은 체계화된 양적 연구 체계를 갖게 되며, 개체계량학과 텍스트 마이닝은 새로운 분석 데이터를 갖추게 되면서 향후 디지털 인문학 연구 활성화의 촉매가 될 수 있다고 보았다. 본 연구에서는 역사학 데이터를 대상으로 텍스트마이닝과 개체계량학의 방법론을 활용하고자 하였다. 분석 대상으로는 임진왜란 시기의 사료 데이터인 ‘조선왕조실록(朝鮮王朝實錄)’의 ‘선조실록(宣祖昭敬大王實錄)’, ‘선조수정실록(宣祖修正實

錄)’과 ‘난중잡록(亂中雜錄)’, ‘징비록(懲毖錄)’으로 설정하였다. 임진왜란은 인문학 연구 중에서 대중적 관심과 학계의 지속적인 관심을 바탕으로 하여 새로운 연구들이 나타나며 성장세를 보이는 주제이다(조인희, 2020).

2. 선행 연구

2.1 임진왜란 연구

임진왜란은 16세기 아시아 지역에 명(明)을 중심으로 형성된 조공책봉(助工冊封) 체제 붕괴에 영향을 미쳤으며, 이는 새로운 국가 관계 형성으로 이어졌다. 구체적으로 임진왜란이 미친 영향을 살펴보면 먼저 명은 만력삼대정(萬曆三大征)으로 불리는 세 개의 큰 전쟁을 겪으며 많은 국력을 소모하였다. 특히 임진왜란은 앞선 두 전쟁보다 많은 재정을 지출했으며, 이러한 혼란을 노린 변방 신흥세력 성장으로 내정이 혼란해졌다(하우봉, 2011). 조선은 전란으로 인한 피해 복구에 국력을 다하였으나, 국가 기반 약화와 함께 양반 사족의 지배체제가 강화되었다(김현영, 2012). 일본은 임진왜란을 기점으로 정치적으로 큰 요동을 보였다. 정유재란 시기에 태합(太閤)인 도요토미 히데요시(豊臣秀吉, 1537-1598)의 사망으로 인하여 다이묘(大名)들이 이분(二分)되며 내부적으로 큰 혼란이 발생했다(이계황, 2015).

임진왜란은 동아시아 삼국에 정치 변화를 넘어서 여러 부문의 변곡점으로 작용하였기에 많은 연구가 수행되었으나 그간의 성과에 대하여 정리한 연구사 연구는 최근에서야 이루어졌다.

근래의 임진왜란 연구는 '동아시아 전쟁'으로서의 인식이 제안되며 이에 관한 연구가 중심을 이루고 있다(하우봉, 2011). 본 연구는 임진왜란의 전개와 결과보다는 임진왜란에 관한 연구사 연구와 외교와 군사 부분에서 중책을 수행한 인물에 관한 연구를 정리하였다. 이러한 정리를 바탕으로 본 연구의 계량 분석으로 기존 연구의 보완점을 파악과 새로운 결과를 해석하기 위하여 기술하였다.

노영구(2010)와 이계황(2009)은 국내 연구사 전반을 정리하였다. 노영구(2010)는 통시적인 시각을 바탕으로 1945년의 해방 이후부터 2000년대까지 국내 임진왜란 연구를 살펴며 주제적 변화를 분석했다. 연구 주제 변화가 사회 변화와 밀접함을 입증하며 근래 동아시아 국제전쟁으로서 임진왜란의 인식을 강조하였다. 이계황(2009)은 해방 이후부터 2000년대 초반의 임진왜란 연구사를 정리하며, 단행본과 학술지로 구분하여 학술지는 총 8개 분류(정치, 경제, 사회, 인물, 의병, 전쟁, 외교, 기타)로 나누어서 총 729개의 문헌에 대해 분류하는 통계적인 방법으로 접근함으로써 임진왜란 연구에 내재한 문제를 지적하였다.

나카노 히토시(2009)는 일본 임진왜란 연구사 분석 시대를 17세기부터 21세기 초반까지 정리하였다. 처음으로 '동아시아 전쟁'의 관점을 제시한 李啓煌(1997)의 연구를 거론하며, 동아시아 국제전쟁으로 임진왜란에 대한 접근을 통해 조선과 명, 일본 관계에 관해서 입체적인 연구의 필요성을 주목하였다. 조인희(2020)는 정유재란에 대한 연구사를 정리하여 정유재란 시기 일본의 전략적 목표, 점령 방식 변화를 바탕으로 재침략 원인과 정유재란 시기 전환

변화, 일본의 목표 변경 사유를 찾고자 하였다.

본 연구는 인물 개체로는 김성일(金誠一, 1538-1593)과 이항복(李恒福, 1556-1618)에게 집중하였다. 두 인물은 임진왜란 시기에 외교와 군사 영역에서 중요한 구실을 하였으나 그들에 관한 연구는 적은 수만 찾을 수 있었다. 김성일이 주제인 연구는 2개, 이항복을 주제로 한 연구는 1개가 확인되었다. 김성일은 일본과의 외교 전반을 담당하였으며 임진왜란 발발 이후에는 경상도 지역을 방어하였다(하우봉, 2012). 이항복은 정여립 모반 사건(己丑獄事)을 처리하였으며, 임진왜란 발발 이후에는 선조(宣祖, 1552-1608)의 피난길을 호종(扈從)하였고 명과의 외교 전반을 담당하였다(한국학중앙연구원, 2009).

김성일을 제재로 한 연구를 먼저 살펴보면, 하우봉(2012)은 김성일의 통신사 귀국 보고와 관련하여 당대 조선 지식인들의 일본 인식과 결부하였고, 김학수(2014)는 김성일의 인적 네트워크를 탐구하였다. 하우봉(2012)은 통설(通說)이었던 김성일의 오판에 대한 비판 대신에 새로운 시각을 제시하였다. 일본에 대한 오판을 당시 조선인들이 일본에 대해서 가지고 있던 통념(通念)을 원인으로 보며 결정권자였던 선조와 조정대신(朝廷大臣)의 책임론을 제기하였다. 김학수(2014)는 임진왜란 시기에 김성일이 활용한 인적 네트워크에 대한 검토와 병마절도사로서 호남지방을 방어한 공로를 바탕으로 김성일을 재평가하는 견해를 제시하였다. 또한, 김성일이 비록 일본의 군세(軍勢)와 침략 정황에 대해서 오판하였지만, 왜란 발발 이후에는 경상도를 중심으로 방어와 민심을 추스른 활동을 높이 평가하였다. 이항복에 관한 연구는 주로 그의 문학적 업적에 초점이 맞추어져 있었

다. 이항복의 문학적 소양과 작품과 관련된 연구는 다수가 존재했으나 선행 연구 중에서 이항복의 외교적 능력을 다루었던 연구는 정억기(2007)의 이항복 외교활동의 성격에 대한 분석만을 찾을 수 있었다. 정억기(2007)는 이항복의 외교활동을 임진왜란 발발 직후 명(明)에 원군 요청과 강화교섭 시기의 진주사(陳奏使) 파견, 광해군 시기의 외교활동을 중심으로 분석하고 실익을 증요시한 외교로 평가하였다.

2.2 개체계량학 연구

개체계량학(Entitymetrics)은 Ding et al.(2013)의 연구로 처음 제시되었으며 문헌에 분포된 다양한 개체(Entity)를 활용하여 새로운 지식을 추론하는 방법론이다. 기존 문헌 관련 연구에서 개체였던 저자, 초록, 키워드, 저널, 참고문헌을 평가적 개체(Evaluative-entity)로 보며 키워드, 주제, 핵심 방법론이나 이론을 지식적 개체(Knowledge-entity)로 분류하였다. 개체계량학은 문헌 기반 발견(Literature based discovery), 계량 서지학(Bibliometrics), 텍스트마이닝(Text-mining)의 개념과 방법론이 융합된 학문 분야로, 디지털 인문학이 추구하는 다학제성 연구와도 부합한다고 볼 수 있다.

개체계량학 연구에 대해서 살펴보면, Ding et al.(2013)은 문헌의 구성요소인 개체의 중요성을 주목하며 기존 연구 방법들을 확장하여 개체계량학을 처음으로 제안하였다. 해당 연구는 PubMed와 PubMed Central의 데이터를 대상으로 선정하여 당뇨병 치료제인 메트포르민(Metformin)과 관련된 문헌들의 인용 정보 등의 관련 개체 간 네트워크를 분석하였다. Yu et al.

(2015)은 생의학 전문 데이터베이스가 가지고 있는 영향력 측정을 위하여 데이터베이스 링크 네트워크를 생성하였다. 논문과 참고문헌에서 인용 정보를 추출하여 데이터베이스 링크 네트워크에서 나타나는 경로로 개체 관계를 분석하였다.

국내에서는 함정은과 송민(2015)의 연구로 개체계량학이 처음 소개되었다. Swanson(Don R. Swanson, 1924-2012)이 제안한 ABC 모델과 문헌의 인용 정보를 활용하여 유의미한 관계를 갖는 개체들을 효율적으로 찾는 방법을 제안하였다. PubMed Central의 데이터를 대상으로 기존의 ABC모델과 개체계량학을 적용한 연구 결과를 비교하였다.

박진균, 김택윤, 송민(2017)은 인문학 데이터에 개체계량학을 적용하였다. 윤동주(尹東柱, 1917-1945)를 제제로 한 연구의 초록을 대상으로 인명, 작품명을 개체로 선정하여 관계를 분석하였다. 토픽 모델링과 저자 동시인용 분석을 통해서 연구 동향을 분석과 개체계량학을 이용하여 윤동주 관련 연구가 생애, 시, 비교문학, 번역문학 등 다양한 영역에 걸쳐서 이루어진 특징을 도출하였다.

2.3 다중 네트워크 분석

네트워크 분석은 다양한 대상의 관계 패턴과 구조를 측정할 수 있다. 네트워크 분석 대상을 사회적 관점으로 전환하면 소셜 시스템(Social System)의 구성원 간 관계를 분석하는 소셜 네트워크 분석이 된다(곽기영, 2017). 다양한 데이터 분석에서 사용되는 네트워크 분석 방법의 하나인 다중 네트워크는 분석 데이터 용량이 증

가하면서 대량의 데이터를 조망할 수 있다는 장점을 활용하여 사회학, 생물학, 물리학, 공학 등 다양한 연구에서 이용되고 있다. 하지만 다중 네트워크를 활용한 연구는 증가하였으나, 정의에 관한 연구가 부재하여 다른 네트워크 개념과 혼재하였다. Mikko et al.(2014)는 이러한 작금(昨今)의 개념이 혼재한 상황을 비판하며 다양한 네트워크의 정의를 구분하였다. Mikko et al.(2014)는 '다중 네트워크(Multilayer Network)'에 대해서 두 가지로 구분할 수 있으며, 두 종류 모두 차원 간에 분리가 부재하다고 하였다. 첫 번째 다중 네트워크는 각 차원의 노드 수가 같으며 노드 간의 연결이 정렬된 형태로 분류하였다. 두 번째 다중 네트워크는 계층 간에 노드 연결이 범주화 가능하며 '차원(Aspect)' 간에 층위가 없다고 하였다. 본 연구는 노드 개수와 공통 개체 간 연결이 자유로운 '다중 네트워크(Multi-layer Network)' 분류를 참조하여서 네트워크를 구축하였다.

다중 네트워크는 개체 간 연결 관계를 입체적으로 표현할 수 있다. 이러한 특징으로 의학 분야에서는 다중 네트워크 계층 간 층위에 구분을 부여하는 다계층 네트워크 분석의 비율이 높다. 다계층 네트워크는 이종(Heterogenous), 또는 동종(Homogenous) 네트워크 간 비교 방법으로서 층위가 존재하여 네트워크 간 연결 순서에 따라서 관계 형성에 영향을 줄 수 있다. 이러한 다계층 네트워크를 이용한 사례로는 Berenstein et al.(2016)의 연구가 있다. 해당 연구는 다계층 네트워크의 각 계층을 약물, 단백질, 작용 대상으로 구분하고, 3개의 계층을 가진 네트워크로 공통 개체 간 연결 관계를 표현하였다.

역사학 분야에서는 Vugt(2017)의 다중구조 네트워크를 이용한 17세기 유럽의 소셜 네트워크 분석 연구가 있다. 해당 연구는 우편물을 활용하여 니콜라스 하인시우스(Nicolaas Heinsius, 1620-1680)와 메디치 가문 관계를 분석하였다. 우편물과 공문서, 우편물 송장을 다중 네트워크로 구현하여 피렌체에서 활동했던 안토니오 말리아베키(Antonio Magliabechi, 1633-1714)가 17세기 소셜 네트워크에서 가진 중요성을 도출하였다.

2.4 토픽모델링

토픽 모델링은 비정형 대용량 텍스트 데이터에서 주제를 찾기 위한 분석 방법으로 단어를 문헌을 구성하는 성분으로 보며, 이에 따라서 문헌을 주제를 대표하는 단어 구성체로 가정한다(Blei, Ng, & Jordan, 2003). 그리고 Steyvers, Griffiths(2007)에 따르면 학습대상인 문헌에 대해서 주제 발견, 문헌 식별 및 유사성 산출을 통하여 '주제(Topic)'라는 새로운 데이터를 생성하기에 생성적 모델(Generative model)이라 볼 수 있다. 변천에 대해서 간략하게 살펴보면 대표적으로 벡터 공간 모델(VSM), 잠재 의미 분석(LSI), 확률적 잠재 의미 인덱싱(PLSI), 잠재 디리클레 할당(LDA)이 있으며, 근래의 연구에는 잠재 디리클레 할당이 많이 사용되고 있다(Barde & Bainwad, 2017). LDA 모델링은 주제를 생성하는 모델로서, 앞선 모델들이 문헌에 기반한 유사 데이터로 주제를 수동적으로 부여된 번호에 의존한다는 선행 모델들이 가졌던 한계점을 보완했다(Blei, Ng, & Jordan, 2003). 그렇기에 LDA 모델에서는 문헌별 주

제 생성을 확률적인 분포를 추산하여 타당한 주제 분석이 가능하다. 그리고 LDA 모델에서 발전한 모델로는 매개변수를 추가한 디리클레 다항 회귀(DMR)와 디리클레 다항 회귀에 매개변수를 2개로 확장한 일반화 디리클레 다항 회귀(Generalized DMR)가 있다(Lee & Song, 2020; Mimno & McCallum, 2008).

토픽 모델링 분석은 학술 문헌의 주제 추출, 트렌드 분석에 주로 활용되어왔다. 강범일, 송민, 조화순(2013)은 오피니언 마이닝을 이용하여 신문이 내포하고 있는 시사적인 이슈에 대한 저자의 관점과 의견을 파악하였다. 통시적으로 2012년 대선 직전의 2년간 발행된 신문을 대상으로 토픽 모델링 방법을 사용하여 추출된 주제에 대해 오피니언 마이닝을 수행하였다. 김용환, 김유신(2019)은 국내 학술지 헬스케어 관련 논문 데이터를 수집하여 연구 동향과 연구 영역을 분석하여 주제들의 추세 변화를 분석했다. 남춘호(2016)는 문헌 기반 연구에 사용되던 토픽 모델링을 디지털화된 인문학 자료에 적용하여 활용 가능성을 살폈다. 해당 연구는 경상북도 거주 농민이 30년간 작성한 일기에 대해서 토픽 모델링 분석을 적용하여 인문학 데이터를 분석하는 방법으로써 토픽 모델링의 타당성을 제시하였다.

3. 연구방법

3.1 연구 문제

본 연구에서는 임진왜란 시기 데이터를 대상으로 텍스트 마이닝과 개체계량학 분석 방법을

적용함으로써 향후 디지털 인문학 연구에 있어서 구체적인 방향성을 제안하고자 했다. 텍스트 마이닝과 개체계량학은 대규모 텍스트 자료를 구축하고 있는 인문학에 적절한 데이터 분석 방법이며, 이미 방법론적으로 타당성을 축적해왔기에, 디지털 인문학 데이터 분석에 최적의 방향성을 제시할 수 있다. 이러한 이해를 바탕으로 아래의 세 가지 연구 문제를 설정하였다.

- 연구질문 1. 동시대 사료로 구축된 네트워크에서 어떠한 관계와 구조가 나타나는가?
- 연구질문 2. 시간 흐름에 따라서 어떠한 주제적 특성과 변화를 보이는가?
- 연구질문 3. 계량 분석을 통해서 나타난 개체 관계는 기존 연구와 비교하여 어떠한 차이점을 갖는가?

연구 문제에 대해서 살펴보면, 첫 번째 연구 질문은 다중 네트워크를 활용하여 사료 속의 인물 개체 관계를 파악하고자 하였다. 난중잡록, 징비록, 실록을 기반으로 다중 네트워크를 구축하고, 사료의 종류에 따라서 나타나는 인물 관계의 특징을 찾고자 하였다.

두 번째 연구 질문은 DMR 토픽 모델링을 활용한 실록의 주제 변화 탐구와 연관되어 있다. DMR 토픽 모델링은 메타 데이터로 시간 정보를 수용할 수 있으므로, 이러한 특징을 활용하여 주제 변화에 대해서 시계열적 분석을 할 수 있다. 이에 DMR 토픽 모델링을 이용하여 실록의 임진왜란과 정유재란 기간의 주제 변화를 살피고자 하였다.

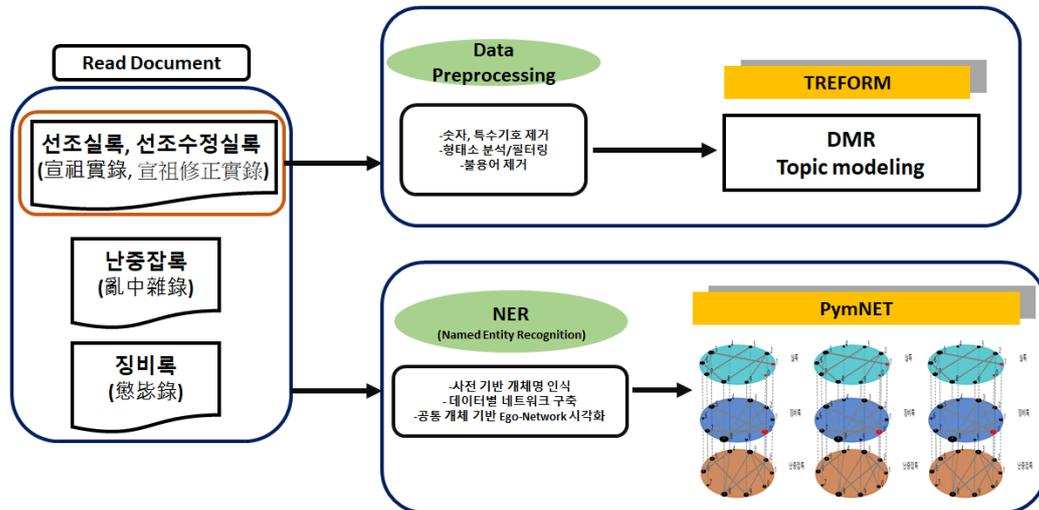
세 번째 연구 질문은 계량 분석을 통해서 나

타난 개체 관계와 기존 연구에서 도출된 개체 관계의 차이를 밝히고자 하였다. 통설적으로 형성된 개체 관계 외에 본 연구에서 시도한 DMR 토픽 모델링에 따른 주제 변화와 다중 네트워크를 기반한 새로운 인물 개체 관계 발견 가능성을 시사하고자 했다.

3.2 연구 개요

본 연구는 역사학 데이터에 대해서 텍스트 마이닝과 개체계량학의 분석을 활용한 연구 과정을 <그림 1>과 같이 수립하였다. 분석 데이터는 총 세 개의 전산화된 사료를 수집하였다. 수

집 대상은 이상훈(2010)이 문체(文體)와 저술주체(主體)를 기반으로 관찬사료(官撰史料), 친필 필사본(筆寫本), 야담(野談)과 야사(野史), 문집(文集), 실기(實記), 일기(日記)로 분류한 임진왜란 사료 기준을 참조하였다. 또한, 모든 사료가 전산화되어 있지는 않기에 전산화된 데이터를 우선시하여 관찬사료, 친필 필사본류, 야담 야사류에서 각각 1개의 사료를 선택하였다. 관찬사료는 조선왕조실록(朝鮮王朝實錄), 친필 필사본은 징비록(懲毖錄), 야사류는 난중잡록(亂中雜錄)을 수집하였다. 수집한 데이터의 통계는 <표 1>과 같으며, 수집한 데이터는 텍스트 파일로 저장하였다.



<그림 1> 연구의 개요

<표 1> 임진왜란 시기 분석 데이터 통계

문헌 종류	문헌 수
조선왕조실록(선조실록, 선조수정실록)	11,506
난중잡록	737
징비록	60

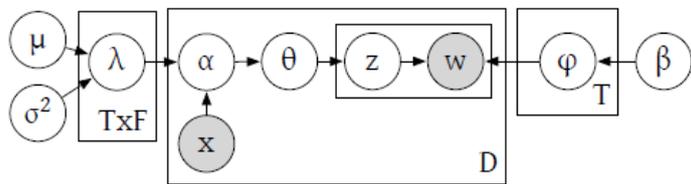
수집된 데이터인 실록, 난중잡록과 징비록은 문헌 수에서 차이가 나타났다. 이 같은 데이터 크기 차이는 사료의 종류에 따라서 사료의 특성, 저술 시기, 저술 근간 자료에 따라서 나타난 것으로 추정되었다. 실록은 사관(史官)에 의해 기록된 방대한 사초(史草)를 바탕으로 저술되나, 난중잡록은 조정남(趙慶男, 1570-1641)이 전란(戰亂) 중에 저술한 일기 형태이며, 징비록은 유성룡(柳成龍, 1542-1607)이 유배(流配) 중에 집필하였다(이상훈, 2010).

수집한 데이터는 분석 용이성을 위하여 메타 데이터로 부여된 날짜 정보를 기준으로 정렬하였다. 실록은 같은 일자에 여러 기사가 있었다. 실록의 서술 특징으로 각 기사를 하나의 행으로 정렬하여 날짜 순서에 따라서 제열(齊列)하였다. 연구에서 사용된 데이터는 원문 사료를 바탕으로 번역된 국문 번역본이다. 데이터 수집 과정에서 전문을 수집하고 영역별 구분자를 삽입하였다. 전처리 작업이 수행된 자료를 수집하였기에 특수기호, URL, 이모티콘에 대한 전처리 과정은 별도로 필요하지 않았다.

분석 방법으로는 다중 네트워크(Multi-layer Network) 분석과 DMR 토픽 모델링을 사용하여 새로운 개체 관계과 주제 변화의 탐색 가능

성을 검토하고자 하였다. 소셜 네트워크 분석과 다중 네트워크 구조를 결합하여 사료에서 추출한 인물 관계를 분석하였다. 데이터로 활용한 사료는 층위가 없기에 다중 네트워크를 활용하여 구현했으며, 각 차원 네트워크는 에고 네트워크(Ego-Network)로 구축했다. 에고 네트워크는 중심인 에고(Ego)와 주변부인 알터(Alter)로 구성되며 이들 간 관계를 탐색할 수 있다(곽기영, 2017). 이러한 에고 네트워크 특징을 적용하여 다중 네트워크로 공통으로 출현하는 개체를 도식화하였다. 개체는 국사편찬위원회에서 제공하는 조선왕조실록 인명사전(National Institute of Korea History, 2021)을 바탕으로 인명, 관직명으로 사전 기반 개체명 인식(Named Entity Recognition)을 수행했다. 네트워크 시각화는 Gephi¹⁾, 다중 네트워크는 Pymnet(Mikko et al., 2014)을 활용하였다.

토픽 모델링 분석으로 본 연구에서는 DMR 토픽 모델을 사용하였으며, <그림 2>는 해당 모델의 연산 과정 도식이다. LDA 모델링과 DMR 모델의 차이점은 연산 과정에서 주제 문헌 밀집도를 결정하는 'α'의 위치 차이이다. DMR 모델은 'α'가 매개변수로 작용하며, 'σ'가 메타 데이터에 따른 주제 분포 편차를 조절하는 하이퍼 파



<그림 2> DMR 모델(Mimno & McCallum, 2008)

1) <https://gephi.org/>

라미터(Hyper Parameter)로 작용한다(Mimno & McCallum, 2008). 이러한 구조로 DMR 토픽 모델링은 연구자가 부여하는 메타 데이터 'σ'에 따른 주제적 변화 추세를 분석할 수 있다. 세부적으로 조정할 수 있는 하이퍼 파라미터인 'α', 'η', 'σ', 'ε'는 프레임워크 기본값으로 실행하였다.

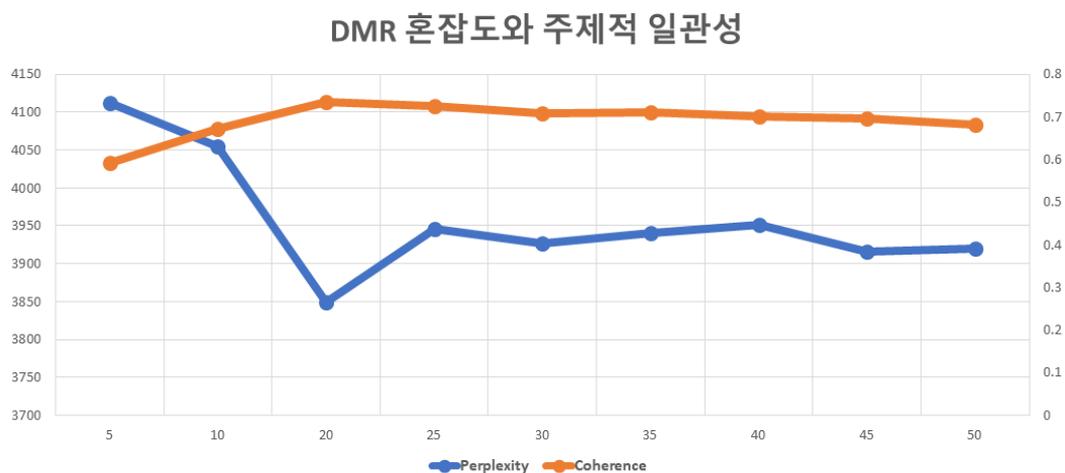
토픽 모델링 분석의 주제 개수인 'K'는 연산 결과에 영향을 준다. 주제 개수는 베이지안(Bayesian) 모델 기반 추산 방법, 혼잡도(Perplexity)를 이용한 방법과 주제적 일관성(Topic Coherence)을 활용함으로써 타당한 주제 개수를 결정할 수 있다. 본 연구에서는 혼잡도와 주제적 일관성을 사용하여 주제 개수를 정하였다.

Steyvers, Griffiths(2007)는 혼잡도를 다양한 언어 모델에 대한 내재적 평가 척도로써 제시했다. 혼잡도는 토픽모델링 분석의 주제 개수를 로그 가능 수치를 활용하는 지표로써, 분석 방법의 타당한 주제 개수인 'K'의 설정 근거로 이용할 수 있다. 혼잡도가 낮은 값이 산출되면 문헌

이 실제로 가진 결과를 반영했다고 볼 수 있다. 주제적 일관성도 타당한 토픽모델링 분석을 위한 주제 개수 설정을 위한 지표다(Newman et al., 2010). 주제적 일관성은 외부에서 구축한 말뭉치를 사용하여 추출된 주제들이 가지는 주제적 일관성을 평가하며, 추출된 주제들이 일관적일수록 높은 수치와 주제적인 유의미성을 의미한다. 본 연구에서는 주제 개수 'K' 값의 결정을 위해서 내재적 평가 지표인 주제적 일관성과 혼잡도를 기준으로 삼고, Tomotopy(Lee, 2021)의 메소드를 사용했다.

주제 개수 'K'를 하이퍼 파라미터로 수집된 조선왕조실록 데이터의 전체 기간을 대상으로 혼잡도와 주제적 일관성을 산출하여 <그림 3>과 같이 정리했다. 이를 바탕으로 적절한 주제 개수 'K'를 두 개의 지표가 겹치는 지점인 'K=20'을 주제 개수로 설정하였다.

본 연구에서는 DMR 토픽 모델링의 시계열적 분석을 조선왕조실록을 대상으로 수행하여 주제 변화를 분석하였다. 토픽 모델링 분석과



<그림 3> DMR 주제 개수에 따른 혼잡도와 주제적 일관성

정에서는 자연어 처리 과정을 파이프라인 형태로 구조화한 프레임워크인 Treform(Song, 2021)을 사용하였다. 전처리 과정에서 DMR 모델링을 위해서 연도를 각각의 데이터 행에 메타 데이터로 첨부하였다. 수집된 데이터의 토큰화에는 NLTK²⁾의 Splitter 메소드를 이용하였고, 형태소 분석과 품사 태깅 등의 전반적인 처리 과정에서는 형태소 분석기인 Mecab(Eunjeon, 2018)을 활용하여 수행하였다. 품사는 형태소 분석기의 필터를 사용하여 명사(NN*) 전반에 대해서 추출하였다. 성능 비교 결과, Mecab은 여타 형태소 분석기보다 빠르며, 정확도와 사용자 사전 용이성이 높다. 다른 형태소 분석기와 성능에 대한 검증한 결과는 <별첨 표 1>과 같다.

4. 연구 분석 및 결과

4.1 다중 네트워크 분석

조선 시대 인물 개체 네트워크에 관한 선행연구가 희소하므로 조선 시대 사료에 대한 네트워크 분석 자체로도 의미를 찾을 수 있었다. 분석 대상으로는 임진왜란과 정유재란 시기에 저술

된 난중잡록, 징비록, 실록을 선정하였다. 인물, 관직으로 구축된 사전을 이용한 개체명 인식 방법을 활용하여 세 개의 사료 데이터에 대해서 소셜 네트워크를 추출하였다. 단순 명사 동시출현 빈도보다는 데이터에서 나타난 인물 관계 파악을 하기 위하여 소셜 네트워크를 구축하였다. 공통 출현 개체를 분석하여 이를 기반으로 예고 네트워크를 생성하였다. 다중 네트워크 분석은 난중잡록, 징비록, 실록을 기반으로 새로운 인물 간 연결 관계를 탐색하고자 하였다. 네트워크 분석을 활용하여 통설에 대한 근거 제시와 사료에 대해서 소셜 네트워크 분석을 적용함으로써 해당 방법의 가능성을 제안하고자 하였다. 전체 네트워크에서 도출된 개체에 대한 통계는 <표 2>와 같다.

난중잡록, 징비록, 실록에서 추출한 약 3,300개의 노드 중에서 세 개 데이터 모두에서 출현하는 59개의 공통 개체를 확인하여 <표 3>으로 정리하였다. 공통 개체 전체에 대한 분석보다는 기존에 조명받지 못했던 개체와 외교 부분에서 중심역할을 수행한 이들을 기준으로 다중 네트워크를 구축하였다. 다중 네트워크 분석에 앞서서 표준화 중심성(Standardized Centrality)을 적용하여 네트워크 규모가 다른 세 개의 네트워크를 개괄적으로 살폈다.

<표 2> 개체명 인식 기반 네트워크 통계

문헌 종류	Node	Edge
조선왕조실록(선조실록, 선조수정실록)	2,807	57,754
난중잡록	406	2,132
징비록	101	440

2) <https://www.nltk.org/>

〈표 3〉 임진왜란 시기 공통 개체에 대한 통계

인물명
경리, 고언백, 광재우, 권율, 권징, 김경로, 김귀영, 김명원, 김성일, 김수, 김응서, 김천일, 낙상지, 동일원, 마귀, 박진, 사대수, 서예원, 석성, 송응창, 신립, 심유경, 양원, 양호, 왕필적, 원균, 유정, 유흥, 윤두수, 윤탁연, 이각, 이광, 이노, 이덕형, 이빈, 이순신, 이양원, 이익기, 이원익, 이일, 이항복, 장령, 정량, 정진, 정철, 조대곤, 조승훈, 참장, 청정, 총병, 최경희, 최홍원, 평수길, 평의지, 평조신, 평행장, 현소, 황혁, 오유충

4.1.1 표준화 중심성

본 연구에서 사용한 데이터들은 사료 유형에 따라서 데이터 크기 차이가 확인되었다. 이러한 데이터 규모 차이는 연산을 통하여 산출되는 중심성 차이와 상관관계를 갖는다. 광기영(2017)은 네트워크 규모에 따라서 상대적으로 노드 간 연결 차이가 발생하며, 네트워크 비교를 위해서는 네트워크 전체 규모라는 영향 요소를 제거한 표준화 중심성을 기반한 비교가 필요하다는 점을 강조하였다. 이에 표준화 중심성을 활용하여 네트워크 크기 영향을 배제하고 개괄적으로 난중잡록, 징비록, 실록에 대해서 바라보고자 하였다. 표준화된 비교 지표로는 표준화 매개 중심성, 표준화 근접 중심성을 사용했다. 두 표준화 중심성은 네트워크 크기에 따른 영

향 요소를 배제함으로써 세 개의 사료 데이터로 구성된 네트워크에서 노드가 가지고 있는 영향력에 대해서 타당한 비교가 가능하였다. 〈표 4〉와 〈표 5〉는 각각 매개중심성, 근접 중심성에 대해서 상위 10개의 표준화 중심성을 가진 개체를 정리하였다. 이와 더불어 표준화하지 않은 각 중심성 상위 개체에 대해서는 〈별첨 표 2〉, 〈별첨 표 3〉, 〈별첨 표 4〉로 정리하였다.

표준화 중심성 결과를 보면, 데이터에서 공통으로 등장하는 개체들을 찾을 수 있었다. 각 중심성에서 나타난 개체들을 분석하면, 표준화 매개 중심성에서는 김명원, 김수가 나타났다. 먼저 김명원(金命元, 1534-1602)을 주제로 한 인물사 연구는 찾기 어려웠다. 다만, 선조실록과 수정 실록에는 그의 행적에 대한 기록이 남

〈표 4〉 표준화 매개 중심성 기반 상위 10개 개체

순위	선조실록·수정실록		난중잡록		징비록	
	개체 / 중심성		개체 / 중심성		개체 / 중심성	
1	총병	0.52	이광	0.20	김명원	0.48
2	유성룡	0.52	판서	0.13	권율	0.19
3	김응남	0.51	광재우	0.11	왕필적	0.13
4	이덕형	0.51	참장	0.08	원균	0.10
5	상서	0.50	청정	0.08	이빈	0.08
6	윤두수	0.50	현령	0.08	이일	0.08
7	김수	0.50	김수	0.07	심유경	0.08
8	경리	0.49	서예원	0.07	지사	0.07
9	김명원	0.49	이복남	0.06	유정	0.07
10	정언	0.48	광영	0.06	사대수	0.05

〈표 5〉 표준화 근접 중심성 기반 상위 10개 개체

순위	선조실록·수정실록		난중잡록		징비록	
	개체 / 중심성		개체 / 중심성		개체 / 중심성	
1	총병	0.09	이광	0.41	김명원	0.52
2	경리	0.06	관서	0.38	권윤	0.42
3	유성룡	0.06	곽재우	0.38	심유경	0.41
4	이덕형	0.06	서예원	0.38	지사	0.41
5	상서	0.05	김수	0.38	이빈	0.40
6	김응남	0.04	정인홍	0.37	이일	0.40
7	윤두수	0.04	김면	0.37	이원익	0.39
8	김수	0.04	현령	0.35	곽재우	0.38
9	관백	0.03	청정	0.35	윤두수	0.37
10	지평	0.03	김천일	0.35	왕필적	0.37

아있었다. 즐기(卒記)를 보면, 성품이 훌륭하며 선조의 피난 시기에는 이원익(李元翼, 1547-1634)과 군을 모아서 왜적에 대비하였으며, 마음이 아름다웠다는 긍정적인 평가가 남아있다.³⁾ 실제로 그는 선조의 피난 이후로 한강 등지에 잔류하면서 왜군으로부터 방어하였으나 임진강 전투에서 패배하였다(한국학중앙연구원, 2009).

표준화 근접 중심성에서는 곽재우(郭再祐, 1552-1617), 윤두수(尹斗壽, 1533-1601), 김수(金睟, 1547-1615)가 공통 개체로 나타났다. 실록과 난중잡록의 표준화 중심성에서 모두 나타난 김수에 대한 인물사 연구는 부재하며 실록에는 민심을 잃은 것에 대한 책망과 성을 버리고 탈주한 사실 등에 대해 남아있었다.⁴⁾ 김수는 난중잡록의 저자인 조경남과 임진왜란 시기의 활동 지역이 겹쳤기에, 높은 중심성을 보인 것으로 추측되었다. 곽재우와 윤두수를 보면, 곽재우는 난중잡록의 저자인 조경남과 마

찬가지로 의병장이었으며 활동 지역이 중첩되었다(한국학중앙연구원, 2009). 윤두수는 임진왜란 이전에 유배되었으나 전쟁 발발 직후 재등용되어 선조의 피난을 호종(扈從)하였으며 정유재란 이후에는 난국(亂國)을 수습하였다(이연숙, 2012). 표준화 중심성에서는 새로운 개체 간 관계의 발견은 어려웠으나 사료의 특성을 반영한 개체 출현과 저자와 활동 지역을 공유한 개체들이 높은 중심성을 갖는 특징을 파악하였다.

4.1.2 사료 개체 기반 다중 에고 네트워크

에고 네트워크는 에고와 알터 간 관계에 집중하여 거대한 네트워크를 미시적(微視的)으로 분석할 수 있다. 특정 관계에 집중할 수 있는 에고 네트워크의 특징을 활용하여 난중잡록, 징비록, 실록의 공통 출현 개체로 1차 에고 네트워크를 구축하였다. 에고 추출 기준은 선행연구를 바탕으로 외교 부분에서 영향력을 보인 조선, 명,

3) 宣祖實錄 157卷 宣祖 35年 丁酉, 宣祖修正實錄 36卷 宣祖 35年 戊午
 4) 宣祖實錄 27卷 宣祖 25年 丙辰, 宣祖實錄 23卷 宣祖 25年 丁亥

일본의 개체를 선정하였다(김경태, 2019; 김학수, 2014; 나카노 히토시, 2009; 노영구, 2010; 이계황, 2009; 허지은, 2004). 조선은 김성일(金誠一, 1538-1593). 명은 석성(石星, ?-1599), 일본은 야나가와 시게노부(平調信, ?-1605)를 선택하였다. 다중 네트워크가 입체적으로 연결 관계를 투사하는 특징을 활용하여 각 사료가 가진 인물 개체 간 연결을 확인하고자 하였다. 다중 네트워크의 토대인 예고 네트워크는 김성일, 석성, 야나가와 시게노부로 추출하였으며, 다중 네트워크의 차원은 난중잡록, 징비록, 실록으로 구성했다.

예고로 선택된 김성일은 선조실록에 근거하면 1589년 11월에 상사(上使)인 황윤길(黃允吉, 1536-?)과 함께 통신사(通信使)로써 일본을 다녀왔다.⁵⁾ 이후 1591년 조선으로 복귀하여 조정에서 황윤길은 병화(兵禍)가 있을 것이라 하나, 이에 대해서 김성일은 이견(異見)을 표하였다. 김성일은 침략과 관련된 정황을 찾기 어려웠으며 황윤길의 의견으로 인해서 민심이 동요될 것을 우려하였다.⁶⁾ 김성일은 임진왜란 발발 이전에도 통신사로 일본을 다녀오는 외교 업무를 맡았으며, 통신부사로 다녀온 경험을 집필한 '해사록(海槎錄)'을 남겼다. 왜란 발발 이후에는 경상도 초유사(招諭使)로 임명되어 경상도 지역 민심 수습과 산개된 관군들을 수습하였다(김경태, 2014; 하우봉, 2012; 한국학중앙연구원, 2009).

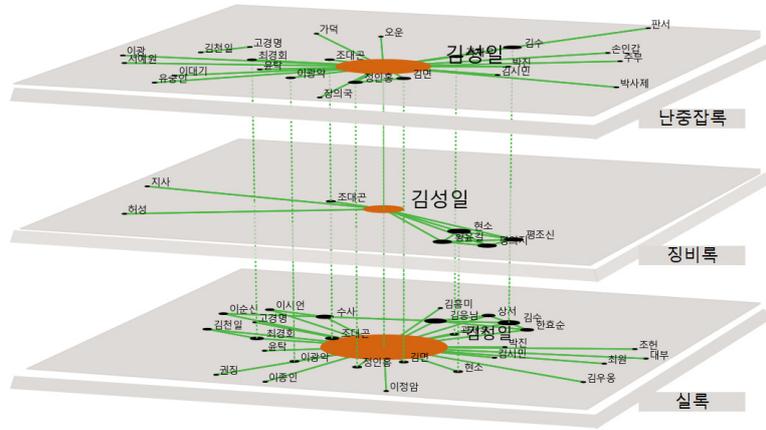
석성과 평조신을 살펴보면, 석성(石星, 1537-1599)은 명의 병부상서(兵部尙書)로서 임진

왜란 시기 군 지휘권을 맡고 있었다(한국학중앙연구원, 2009). 임진왜란 직후 석성은 조선에 대해 원군(援軍) 파병을 적극적으로 추진했으나, 강화교섭 시기부터는 강화파(講和派)로 앞장서며 일본과의 강화교섭을 위해서 심유경(沈惟敬, 1537-1597)을 파견하였다(김경태, 2014; 허지은, 2004). 야나가와 시게노부(柳川調信, ?-1605)는 실록에서는 평조신(平朝臣)으로 기록되어 있으며, 대마도(對馬島)의 다이묘인 소요시토시(宗義智, 1568-1615)의 가신(家臣)으로 임진왜란 발발 이전부터 조선과의 외교를 전담하였다(김경태, 2014; 하우봉, 2012). 예고로 선택한 개체들은 외교 부문에서 중요한 역할을 했던 이들이다. 다만, 석성은 김성일과 평조신처럼 외교 관계보다 명군의 파견과 같은 군사적 영역에 영향력이 치중되어 있으나, 심유경의 파견과 강화교섭과 관련한 허위 보고 사건 등 중요한 외교 안전과도 연관성을 가지고 있었다. 다양한 평가가 가능하거나 기존 연구에서 중요하게 여겨지지 않았던 이들을 중심으로 관찰함으로써 다중 네트워크 분석에 의의를 부여할 수 있으리라 생각된다.

김성일 예고 네트워크를 다중으로 시각화한 <그림 4>를 보면 세 개 네트워크에서 공통으로 조대곤(曹大坤, 생몰미상)이 나타났다. 조대곤은 임진왜란 발발 당시에 경상도 병마절도사(兵馬節度使)로 있었으며, 김성일은 전쟁에 대한 오관으로 인해서 선조와 서인(西人)의 비판을 받았으며, 조대곤의 후임으로 경상도 병마절도사로 파견되었다(김경태, 2014; 하우봉, 2012).

5) 宣祖實錄 23卷 宣祖 22年 壬戌

6) 宣祖修正實錄 25卷 宣祖 24年 丁酉



〈그림 4〉 김성일 다중 네트워크

〈표 6〉 김성일 예고 네트워크 공통 개체

데이터	인물명
선조실록·수정실록, 난중잡록	최경희, 김면, 정인홍, 광재우, 김수, 조대곤, 이광악
난중잡록, 정비록	조대곤
선조실록·수정실록, 정비록	현소, 조대곤
세 문헌 전체 공통 등장	조대곤

하우봉(2012)은 김성일이 임진왜란 초기에 영남지방의 민심을 수습하였으며, 관군과 의병 사이의 갈등을 조율하였기에 일본군으로부터 방어가 가능하였음을 강조했다. 이와 더불어서 김성일이 조정에서 임진왜란에 대한 준비 미흡 문제에 대한 희생양으로써 책임졌던 것으로 분석하였다. 김학수(2014)는 영남지역 방어에 앞장선 김성일이 조정 관료로서 형성한 네트워크와 남명학과(南冥學派)와의 인적 관계를 활용한 능력을 높게 평가하였다.

석성의 예고 네트워크는 인물 관계가 실록에 편중되어 나타나는 경향을 보였다. 실록을 바탕으로 구축된 석성의 네트워크는 다양한 알터 간 관계를 보이나, 난중잡록과 정비록에서는 상대적으로 적은 수의 관계가 확인되었다. 앞선 김성

일의 예고 네트워크와 비교해보면 석성 예고 네트워크는 공통 출현 노드에서 차이를 보였다. 석성 예고 네트워크가 실록에 편향된 관계를 보이는 현상은 실록이 관찬사료(官撰史料)이기 때문으로 보였다. 난중잡록과 정비록은 개인이 저술했으나, 실록은 사초(史草)를 기반으로 실록청(實錄廳)이 편찬하였다. 실록에는 조정에서 거론된 내용과 보고서에 관한 기록되었기에, 석성이 직접 조선으로 파병된 원군을 지휘하거나 저자와의 교류가 부재해도 다양한 알터와의 연결 관계가 나타난 것으로 추측되었다.

난중잡록과 실록에서 석성은 이항복(李恒福, 1556-1618)과 연결되고 있지만, 직접적인 교류 기록은 부재하다. 하지만 이항복이 정응태 무고사건으로 인해 변무사(辯誣使)로서 명에 파

로 기록되었다. 그는 쓰시마(對馬島)의 승려로 임진왜란 시기에는 조선과 강화교섭을 진행하였으며, 이후에는 조선과의 국교 회복과 통신사 재개를 위해서 활동하였다(김경태, 2019; 이계황, 2013). 조선과의 외교 부문에 있어서 중책을 맡았던 겐소는 김성일과는 임진왜란 이전 통신사 파견 중에도 직접적인 교류가 있었다(방기철, 2004). 또한, 김성일은 왜사(倭使) 자격으로 겐소가 한성에 왔던 시기에 대면하였던 적이 있다.⁷⁾ 상대적으로 두 인물 개체 간 관계를 제재로 다룬 연구는 찾기 어렵다. 실록과 징비록에서 겐소가 공통으로 김성일의 알터로 출현하는 모습을 바탕으로 추후 두 개체 간 관계 연구 필요성과 계량적 분석으로 저자 경험과 인물 관계에 대한 도출 가능성을 제기할 수 있었다.

석성의 예고 네트워크는 김성일과 야나가와 시게노부의 예고 네트워크와 비교하면 상대적으로 연결 관계가 실록에 편향된 모습을 보였다. 또한, 새로운 관계로는 이항복과의 간접적인 연결 관계를 확인할 수 있었다. 석성은 앞선 김성일의 예고 네트워크보다는 적은 알터 간 연결 관계를 보였다. 석성은 실록과 난중잡록에서 공통으로 이항복과 개체와 연결되었다. 실록에서 석성은 이항복 이외의 조선 조정 대신과도 연결 관계를 보였으며, 명 관련 개체인 이여송(李如松, 1549-1598), 심유경(沈惟敬, 1527-1597), 오유충(吳惟忠, 생몰미상)과도 알터 관계를 보였다. 석성과 연결된 알터 중에서 가토 기요마사(加藤清正, 1562-1611)가 확인되었다. 허지은(2004)의 연구에 따르면, 정유재란 이후로 명

조정은 강화파(講和派)와 주전파(主戰派)로 이분되었다. 석성은 강화파에 속하였으며 이들의 교섭대상은 고니시 유키나가였다. 반면에 주전파는 가토 기요마사와 교섭을 진행하였다. 이러한 관계를 바탕으로 보면 석성은 고니시 유키나가와 관계가 예상되었으나 네트워크에서는 가토 기요마사와 관계가 형성된 모습을 보였다.

야나가와 시게노부 네트워크는 앞선 김성일과 석성의 네트워크보다 일본과 관련된 개체가 많이 등장하고 있다. 또한, 앞서 석성과 알터 관계였던 이항복이 야나가와 시게노부의 실록과 난중잡록 네트워크에서 공통으로 출현하였다. 석성과 야나가와 시게노부의 네트워크에서 공통으로 출현하는 개체인 이항복이 외교 부문에서 가졌던 영향력의 중요성을 기늬할 수 있었다. 이처럼 각 네트워크에 대한 분석 결과를 토대로 새로운 관계에 대한 탐지와 추후 '명실록(明實錄)', '명사(明史)' 등의 조선 외의 기록을 활용한 연구 필요성을 확인할 수 있었다.

4.2 토픽 모델링 분석

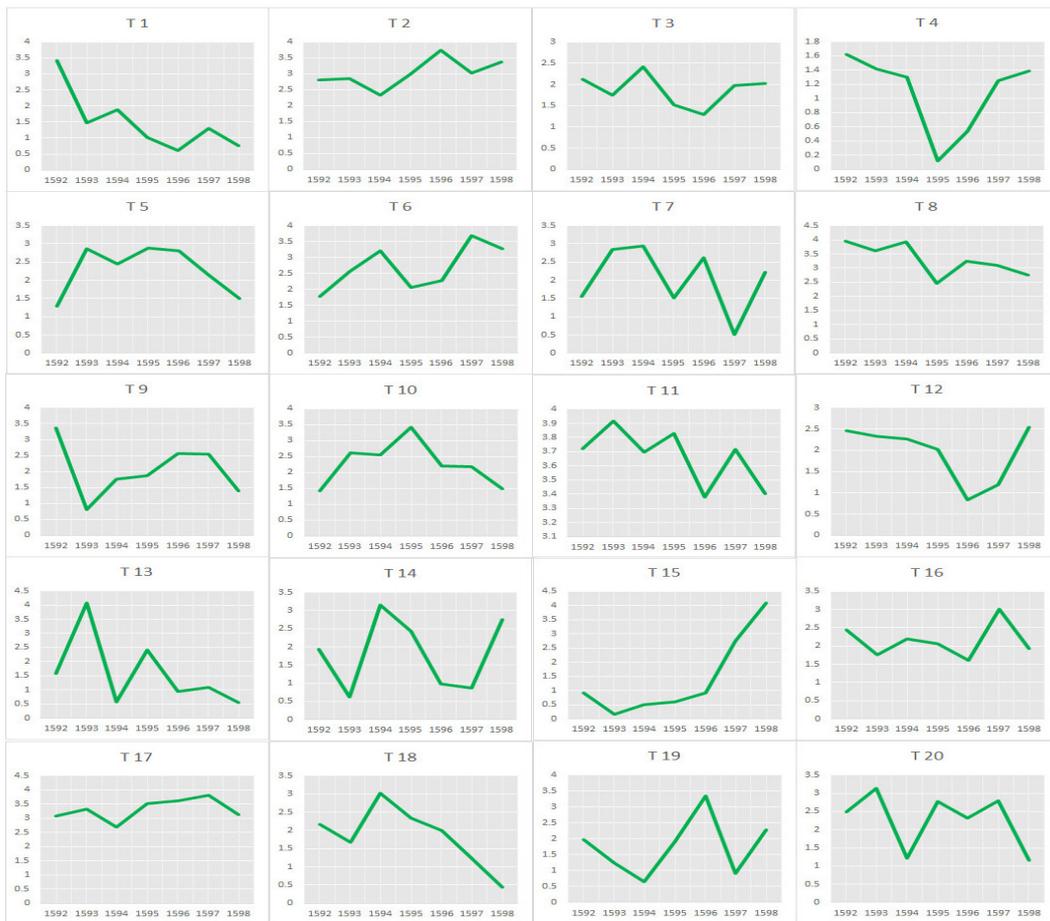
토픽 모델링은 분석 데이터의 시간 변화에 따른 주제적 추세를 찾기 위해 사용하였다. 이에 저술 시기에 대한 메타 데이터를 가지고 있는 실록을 대상으로 하여 시간별 주제 변화를 분석했다. DMR 토픽 모델링은 시간적인 정보를 메타 데이터로 이용하기에 정확한 정보가 있는 조선왕조실록에 대해서만 분석을 수행하였다. 11,506개의 기사를 대상으로 하여서 메타

7) 宣祖修正實錄 25卷 宣祖 24年 丁酉

데이터 범위는 임진왜란이 발발한 1592년부터 일본군이 조선에서 철수한 1598년까지로 설정하였으며, 주제 개수는 앞선 <그림 3>의 결과를 참조하여 혼잡도와 주제적 일관성의 평균치가 낮았던 20개로 설정하였다.

조선왕조실록에 대해 토픽 모델링을 수행한 결과 <그림 7>과 같이 시간 변화에 따라서 주제의 변화를 보였다. <그림 7>의 X축은 시간을 의미하며 Y축은 해당 주제가 X축 시기에 발현할 확률을 의미한다. 주제의 변화 패턴을 찾기

위한 기준으로는 조인희(2020)가 설정한 '명일교섭기(明日交涉期)'를 선택하였다. '명일교섭기'는 1593년 4월 일본군이 한양에서 철수하여 1596년 9월 명의 책봉사(冊封使)가 일본으로 파견되는 약 3년간의 기간을 지칭하며 이 시기에는 강화교섭이 진행되어 전쟁이 소강 국면으로 접어들었다. 임진왜란과 정유재란을 구분하는 기점이며 동시에 전쟁의 양상이 바뀌는 시점이므로 해당 기점으로 주제 변화를 분석하고자 하였다.



<그림 7> 실록 토픽의 시계열적 주제 변화

토픽 모델링으로 추출된 20가지의 주제를 바탕으로 몇 가지의 주제 변화 패턴을 파악할 수 있었다. 주제 2, 17번은 1592년부터 1598년까지 큰 변화를 보이지 않지만, 전체 기간에 걸쳐서 높은 주제 발현확률을 보였다. 각 ‘주제(Topic)’를 구성하는 주제와 주제 단어를 살펴보면 주제 2번은 제례(祭禮)와 관련된 주제로 확인되었다. 주제로는 종묘(宗廟), 태묘(太廟), 예관(禮官)이 나타났으며 주제 단어로는 거행(舉行), 제향(祭享), 사직(社稷) 등의 예조(禮曹)와 관련된 단어 중심으로 나타났다. 주제 17번은 북방의 신흥세력이었던 후금(後金)과 연관된 주제로 보이며, 이는 주제인 만포(滿浦), 평야, 몽고(蒙古)와 주제 단어에서 오랑캐, 호인(胡人), 노을가적(老乙可赤, 1559-1626)의 출현을 바탕으로 알 수 있었다. 17번 주제는 임진왜란으로 약해진 명의 틈을 타서 여진족을 통합하여 후금이 부상하는 상황을 방증하는 것으로 볼 수 있었다.

주제 13과 20은 ‘명일교섭기’를 기준으로 주제 발현확률이 감소하였다. 주제 13은 임진왜란 초기에 명(明)의 원군으로 파병된 장수들과 관련된 주제로 오유충(吳惟忠, 생몰 미상), 군사(軍士), 낙상지(駱尙志, 생몰 미상)가 출현하였으며, 주제 단어는 제독(提督), 경략(經略), 평양, 강화(講和)가 나타났다. 오유충과 낙상지는 명의 장수들로 1593년에 강화교섭이 개시되며 명으로 귀환하였다(한명기, 2002). 주제 20은 임진왜란 초기에 왜군의 포로가 되었던 황정욱(黃廷彧, 1532-1607)과 연관된 주제로 사료되었다. 주제로는 왕법(王法), 형벌, 석방이 나타났으며, 주제 단어로는 죄, 황정욱, 법, 용서, 황혁(黃赫, 1551-1612) 등을 확인할 수 있었

다. 황정욱은 왜군의 포로가 되었던 시기에 아들인 황혁이 작성한 항복 권유문이 빌미가 되어 1593년에 왜군으로부터 석방된 이후에 유배되었다(한국학중앙연구원, 2009). 주제 13과 20의 주제확률 변화는 실록에 대한 시계열 변화와 역사적 사건의 발생 시간이 중첩되는 모습을 예증하였다.

주제 14와 19는 ‘명일교섭기’를 기점으로 주제 발현확률이 증가하는 양상을 보였다. 주제 14는 기축옥사(己丑獄事)와 관련된 주제로 확인되었다. 주제로는 간신, 모함, 정여립(鄭汝立, 1546-1589)이 나타났으며, 주제 단어로는 정철(鄭澈, 1536-1594), 최영경(崔永慶, 1529-1590), 모함, 조정 등을 볼 수 있었다. 기축옥사는 1589년 정여립이 역모를 꾀하였다는 모함을 받아 서인(西人)인 정철을 중심으로 동인(東人)의 인사들을 탄압한 사건이다(한국학중앙연구원, 2009). 주제 19는 주제로 강화교섭과 관련된 주제로 볼 수 있다. 주제로 통신사, 박의검(朴義儉, 생몰미상), 책사(策士)가 출현하였으며, 주제 단어로는 사신, 바다, 배신, 의논이 나타났다. 박의검은 역관(譯官)으로 강화교섭 시기에 명의 진영과 왜군의 진영을 오가며 강화교섭 등의 외교 실무를 담당하였다(한국학중앙연구원, 2009). 주제 14의 경우에는 임진왜란 시기 이전에 발생하였으나 해당 사건에 대한 발현확률이 ‘명일교섭기’ 이후로 증가하였다. 다만 주제 14와 19는 공통적으로 1596년 이후로 주제의 발현확률이 감소하였다.

‘명일교섭기’ 이후에 급격한 증가 추세를 보이는 주제는 주제 4, 15, 18이었다. 주제 4는 정유재란과 관련된 주제로 추측되었다. 주제로는 돌진, 승세, 화전을 보였으며, 주제 단어로는 왜

적, 성, 공격, 사다리, 진격 등이 나타났다. 1594년을 기점으로 급격하게 감소하며, 1596년부터는 반대로 가파른 증가 추세를 보였다. 주제 15는 정응태 무고사건(丁應泰誣告事件)과 관련된 주제로 추측되었다. 주제로는 울산, 진린(陳璘, 1543-1607), 이덕형(李德馨, 1561-1613)이 나타났으며, 주제 단어로는 경리(經理), 제독, 접반사(接伴使)가 출현하였다. 정응태 무고사건은 정유재란 중인 1597년 말에 발생한 울산성(蔚山城) 전투에 대한 피해와 과실 조사를 위해 명 조정에서 파견된 정응태(丁應泰, ?-?)가 조선이 명 장수들의 과실을 변호한 것을 빌미 삼아서 조선과 왜군의 동모(同謀)를 포함한 사건이다(이계황, 1995; 허지은, 2004). 주제 18

은 군사 훈련과 연관된 주제로 파악되었다. 주제로 군사(軍士), 조총(鳥銃), 살수(殺手)가 확인되었으며, 주제 단어로 훈련, 기예(技藝), 병조(兵曹)가 나타났다. ‘명일교섭기’ 이후 시점으로 정유재란이 발생하며 전투에 대한 주제와 병사 훈련에 대한 주제의 증가를 확인하였다. 또한, 명과 조선의 외교적 전쟁으로 강화교섭 체결을 약 6개월 늦춘 정응태 무고사건에 대한 주제가 발생 시점과 비슷하게 증가하는 양상을 확인하였다.

마지막으로 주제 1은 1592년부터 1598년 기간에 지속적인 내림세를 보였다. 주제로는 조헌(趙憲, 1544-1592), 광재우, 김성일이 확인되었고, 주제 단어로는 의병(義兵), 병사, 성, 진주

〈표 9〉 실록, 수정실록에 대한 DMR 토픽 모델링 결과

순번	주제	주제 단어
1	조헌, 광재우, 김성일	의병, 병사, 성, 진주, 장수, 권율, 도망, 김천일, 이빈, 순찰사
2	종묘, 태묘, 예관	제사, 예조, 거행, 제향, 전교, 대신, 뜻, 의논, 사직, 환도
3	군공, 공로, 면천	공, 제수, 전교, 수급, 비변사, 공로, 군공, 벼슬, 품, 관직
4	돌진, 승세, 화전	왜적, 성, 공격, 장수, 사다리, 진격, 군, 주둔, 군대, 형세
5	백성 보호, 백성 원망, 부역	백성, 고을, 폐단, 수령, 원망, 폐단, 감사, 뜻, 국가, 군역
6	관백, 김응서, 우호	청정, 행장, 조선, 일본, 왜인, 관백, 왕자, 강화, 항복, 편지
7	육운, 쌀, 해운	군량, 석, 운반, 곡식, 식량, 쌀, 양식, 수송, 군, 양곡
8	천창, 달무리, 유성	밤, 객성, 동쪽, 천창, 동쪽, 사방, 적색, 일경, 성, 구름
9	시강원, 부응교, 세자	관서, 병조, 참관, 이조, 제수, 홍문관, 사헌부, 사간원, 좌랑, 호조
10	행장, 현소, 조선 왕자	행장, 유격, 노야, 사신, 조선, 일본, 철수, 심유경, 천사
11	황제, 총독, 책봉	동궁, 세자, 칙서, 황제, 뜻, 사신, 책봉, 의논, 군무, 성지
12	전세정, 오유충, 이어백	총병, 유격, 도작, 참장, 출발, 담, 오동, 관서, 병마, 평양
13	오유충, 군사, 낙상지	제독, 경략, 평양, 왜적, 뜻, 임금, 강화, 전진, 진주
14	간신, 모함, 정여립	정철, 죄, 공론, 역변, 영경, 모함, 조정, 무리, 상소, 최영경
15	울산, 진린, 이덕형	경리, 제독, 아문, 접반사, 국왕, 어사, 도독, 주사, 노야
16	영등포, 장문포, 안골포	왜적, 수군, 도원수, 이순신, 원군, 주사, 경상, 장수, 치계, 경주
17	만포, 평야, 몽고	오랑캐, 호인, 노추, 노을가적, 병사, 만포, 부락, 담, 방비, 탈자
18	군사, 조총, 살수	훈련, 기예, 병조, 포수, 법, 시험, 전교, 교련, 장수, 병사
19	통신사, 박의견, 책사	사신, 바다, 배신, 의논, 부사, 뜻, 정사, 황신, 자문, 통신사
20	왕법, 형벌, 석방	죄, 황정욱, 군율, 법, 율, 담, 국문, 용서, 황혁, 도망

(晋州), 장수(將帥)가 나타났다. 의병 활동은 임진왜란 시기에 계속해서 있었기에 발현확률이 지속해서 상위권을 차지하리라 예상하였으나 반대로 발현확률이 감소하는 모습이 확인되었다. ‘명일교섭기’에 일시적으로 전투가 소강상태에 빠지게 되며 내립세를 보일 수는 있으나 전체적인 주제 발현확률이 계속해서 떨어지는 양상에 관해서는 추후 연구를 제안할 수 있다.

문헌을 구성하는 주제와 주제 단어의 시계열 분석을 통하여 시기별 화두인 주제와 언급 빈도가 내립세를 보인 주제를 확인하였다. DMR 토픽 모델링을 통해서 연구자가 사료에 대해서 거리두기를 적용한 거시적(巨視的) 관점으로 문헌을 분석하였다. ‘명일교섭기’를 기점으로 주제들의 패턴을 분석한 결과로 다수 주제가 메타데이터인 시간과 겹치는 주제 변화를 보였으나, 사건이 종료되었음에도 증가세를 보인 기축옥사와 지속적인 내립세를 보인 의병에 대해서는 추가 분석을 제안할 수 있다.

5. 결론

본 연구에서는 난중잡록, 징비록, 실록을 대상으로 텍스트마이닝과 개체계량학을 활용함으로써 기존의 동향 분석 중심의 연구 경향을 넘어서는 디지털 인문학 연구의 가능성을 제시하고자 하였다. 동시대에 저술된 사료를 활용하여 새로운 개체 간 연결 관계와 시간 흐름에 따른 사료 주제 변화를 찾기 위해서 다중 네트워크를 활용한 분석과 DMR 토픽 모델링을 활용하였으며 연구 결과를 요약하면 다음과 같다.

난중잡록, 징비록, 실록을 표준화 중심성으로

분석한 결과를 통하여 공통으로 등장하는 개체를 확인할 수 있었다. 표준화 매개 중심성에서 실록과 난중잡록의 공통 개체로는 김수가 나타났다. 징비록과 실록의 공통 개체로는 김명원이 출현하였다. 표준화 근접 중심성에서는 실록과 난중잡록의 공통 개체로 김수가 확인되었고, 난중잡록과 징비록에서는 광재우가 나타났다. 실록과 징비록은 윤두수가 출현하였다. 결과적으로 윤두수와 김수를 제외한 개체들은 의병장으로서 경상도에서 활동한 공통점을 가진 인물 개체들의 출현을 확인하였다.

김성일, 석성, 야나가와 시게노부의 예고 네트워크를 다중 네트워크로 시각화한 분석 결과로 새로운 연결 관계를 도출할 수 있었다. 다중 네트워크에서 새로운 개체 간 연결 관계로는 김성일은 겐소, 석성은 가토 기요마사, 야나가와 시게노부는 이항복과의 연결 관계가 확인되었다. 각 네트워크의 특징으로는 김성일 네트워크는 실록과 난중잡록의 공통 개체로 의병장과 경상도 지역의 관료들이 다수 등장하였으며, 석성과 야나가와 시게노부는 각 개체의 국가와 연관이 있는 인물들이 중심으로 출현하였다.

김성일의 네트워크에서는 실록과 징비록에서 알티로 등장한 겐소와의 관계를 분석하는 연구를 제안할 수 있었다. 선행연구에서도 해당 관계에 대해서는 크게 주목하지 않았으나 두 개체는 양국에서 외교적으로 중책을 맡았었다. 이를 통해서 조선과 일본의 외교 부문에서 중책을 맡았던 개체 간의 새로운 관계 가능성을 제시할 수 있었다.

석성과 평조신의 네트워크에서는 공통으로 관계를 갖는 개체로 이항복이 도출되었는데, 이로 미루어 보았을 때 이항복의 외교적 능력을 재고

할 수 있었다. 이항복과 석성, 야나가와 시게노부 개체 간 관계에 관한 연구는 찾기 어려우나 두 개체와 직접적인 교류 가능성에 대해서는 추측해볼 수 있었다. 이항복이 두 개체와의 접점을 짐작해보면 석성과는 진주변무사(陳奏辨誣使)로 정응태 무고사건으로 파견되었을 때, 야나가와 시게노부와는 한양에 그가 일본 사신으로 왔던 시기로 추측할 수 있었다. 간접적 관계로 해당 개체 간에는 직접적인 만남이나 교류가 부재했으나 강화교섭과 관련된 사료를 통해서 향후 재해석의 필요성을 제기할 수 있었다. 이처럼 발견된 새로운 관계들을 축대로 명과 일본의 사료를 조선의 사료와 융합한 연구의 필요성을 타진할 수 있다.

인문학 데이터에 시계열 분석의 적용 가능성을 확인하기 위하여 DMR 토픽 모델링을 활용하였다. 해당 분석을 통하여 시간과 주제 간에 상관관계를 부여하여 변화와 특징을 관찰하고자 하였다. 토픽 모델링을 ‘명일교섭기’를 기준으로 분류한 결과 20개의 주제에 대해서 증감의 변화추이 패턴을 보이는 주제들을 도출할 수 있었다. 강화교섭을 기준으로 명의 원군의 귀환과 황정욱의 유배 사건이 시간 변화에 따라서 주제 발현확률이 변화하는 패턴을 보였다. 또한, 정응태 무고사건이 발생한 시점에 맞물려서 주제 발현확률의 증가를 확인하였다.

토픽 모델링 결과로는 지속해서 높은 주제 발현확률을 보인 후금(後金) 주제, 사건 종결 이후에도 높은 주제확률을 보인 정여립 모반사건, 지속적인 내림세를 보인 의병 주제에 관한 추가 연구 필요성을 제안할 수 있었다. 후금 주제는

임진왜란을 축대로 한 동아시아 질서의 변경과도 연관이 있으며, 실록에서 계량 분석을 통하여 해당 주제의 출현은 명(明)의 사료를 이용한 주제 분석 연구의 필요성을 보여주었다. 또한, 정여립 모반사건은 약 3년여간으로 임진왜란 발발 시기에는 관련된 동인(東人)들은 이미 조정에서 퇴출당하였기에 해당 주제가 발현확률을 유지한 배경에 관한 분석도 가능하리라 보았다. 의병 주제는 이미 많은 선행 연구가 이루어졌으나 실록에서 발현확률이 지속해서 감소하는 연유에 관한 연구의 필요성을 제시할 수 있었다.

본 연구에서 활용한 계량 분석 방법은 대량의 데이터에 대한 처리가 쉬우며, 연구자의 주관적 견해를 배제한 연구 결과를 도출할 수 있었다. 또한, 분석 결과의 전산화는 기존의 인문학이 가졌던 고도화로 인해 대중의 접근이 어려워지는 문제를 프로그래밍을 활용하여 접근장벽을 낮추는 등의 해결책으로도 작용할 수 있으리라 짐작한다. 또한, 토픽 모델링 결과를 바탕으로 다양한 데이터의 주제 비교를 활용한 연구 방법의 확장 가능성을 확인하였다.

마지막으로 본 연구의 한계점은 난중잡록, 징비록, 실록으로 한정된 데이터를 활용한 부분을 들 수 있다. 새로운 개체 관계와 주제적인 변화를 탐지한 의의가 있으나 분석 데이터의 한계로 이에 대한 근거와 추가적인 분석으로는 확장하지 못하였다. 한계점을 보완하기 위해 향후 연구로 명(明)과 일본의 사료와 조선의 다른 사료를 활용하여 분석할 필요가 있을 것이다.

참 고 문 헌

- 강범일, 송민, 조화순 (2013). 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. 한국 문헌정보학회지, 47(4), 315-334. <https://doi.org/10.4275/KSLIS.2013.47.4.315>
- 곽기영 (2017). 소셜네트워크분석. 서울: 청람.
- 권윤경 (2018). 새로운 문필공화국을 향하여: 18세기 프랑스사 연구와 디지털인문학의 사례들. 역사학보, 240, 35-74. <https://doi.org/10.16912/tkhr.2018.12.240.35>
- 김경태 (2014). 임진전쟁기 강화교섭 연구. 박사학위논문, 고려대학교 대학원 한국사학과.
- 김경태 (2019). 임진전쟁기(1594년) 조선군과 일본군의 이면교섭 연구. 한일관계사연구, 61, 125-172. <https://doi.org/10.18496/kjhr.2018.08.61.125>
- 김문자 (2017). 家康 국서문제의 僞造 및 改作과 동아시아. 한일관계사연구, 58, 213-254. <https://doi.org/10.18496/kjhr.2017.12.58.213>
- 김바로 (2014). 해외 디지털인문학 동향. 인문콘텐츠학회, 33, 229-254.
- 김바로 (2021). 인공지능 시대의 사전과 데이터 한국역사인물사전(데이터)을 위한 탐색. 동양학, 84, 219-240. <https://doi.org/10.17320/orient.2021..84.219>
- 김용환, 김유신 (2019). 토픽모델링을 이용한 국내 헬스케어 학술연구 트렌드 분석. 한국웰니스학회지, 14(1), 253-262. <https://doi.org/10.21097/ksw.2019.02.14.1.253>
- 김학수 (2014). 김성일(金誠一)의 임란 중 활동과 인적 네트워크. 남명학연구, 41, 61-98. <https://doi.org/10.14381/NMH.2014.41.03.30.61>
- 김현, 임영상, 김바로 (2016). 디지털인문학 입문. 서울: 휴복스.
- 김현영 (2012). 16세기 동아시아 국제질서의 변동과 왜란 전후 조선사회의 변화. 한국사학사학보, 26, 87-116.
- 나카노 히토시 (2009). 文祿, 慶長の 役 연구의 학설사적 검토. 한일역사공동연구위원회 편. 제2기 한일역사공동연구보고서. 서울: 한일역사공동연구위원회, 237-256.
- 남춘호 (2016). 일기자료 연구에서 토픽모델링 기법의 활용가능성 검토. 비교문화연구, 22(1), 89-135.
- 노영구 (2010). 임진왜란의 학설사적 검토. 한일관계사연구논집 편찬위원회 편. 동아시아 세계와 임진왜란. 서울: 경인문화사, 1-35.
- 문수현 (2018). 독일의 디지털 역사학 현황. 역사학보, 240, 1-33. <https://doi.org/10.16912/tkhr.2018.12.240.1>
- 박은재 (2018). 영국 디지털 역사학의 발전과 현황. 역사학보, 240, 75-98. <https://doi.org/10.16912/tkhr.2018.12.240.75>
- 박진균, 김택윤, 송민 (2017). 텍스트마이닝을 이용한 운동주 연구의 개체계량학적 분석. 한국비블리

- 아학회지, 28(1), 191-207. <https://doi.org/10.14699/kbiblia.2017.28.1.191>
- 방기철 (2004). 학봉 김성일의 대일인식. 통일인문학, 42, 21-37.
- 송인재 (2018). 대만 디지털인문학의 발자취와 진화. 역사학보, 240, 127-155.
<https://doi.org/10.16912/tkhr.2018.12.240.127>
- 유재춘 (1986). 임란후 한일국교 재개와 국서개작에 (國書改作) 관한 연구. 강원사학, 2, 153-354.
- 이계황 (2009) 임진왜란을 보는 눈. 동북아역사재단 편. 역사 속의 한일관계. 서울: 동북아역사재단, 122-143.
- 이계황 (2015). 일본근세사. 서울: 해안.
- 이상훈 (2010). 임진왜란관련 사료해제. 한일관계사연구논집 편찬위원회 편. 동아시아 세계와 임진왜란. 서울: 경인문화사, 195-260.
- 이연숙 (2012). 梧陰 尹斗壽의 정치활동과 임진왜란. 한국인물사연구, 17, 69-108.
- 이재연 (2018). 디지털 시대의 인문학에서 디지털 인문학 시대로 한국문학에서 본 디지털 인문학 연구. 역사학보, 240, 157-190. <https://doi.org/10.16912/tkhr.2018.12.240.157>
- 이주영 (2018). 미국에서의 디지털 역사학 발전과정과 최근의 경향. 역사학보, 240, 99-126.
<https://doi.org/10.16912/tkhr.2018.12.240.99>
- 정영기 (2007). 백사 이항복의 외교활동. 한국인물사연구, 8, 47-72.
- 정유경 (2020). 디지털 인문학 분야의 국내외 연구 동향 분석. 정보관리학회지, 37(2), 311-331.
<https://doi.org/10.3743/KOSIM.2020.37.2.311>
- 정은경 (2021). 디지털 인문학 연구 동향 분석: Digital Humanities 학술대회 논문을 중심으로. 한국 문헌정보학회지, 55(1), 393-413. <https://doi.org/10.4275/KSLIS.2021.55.1.393>
- 조원희 (2017). 시론: 네트워크 분석을 통해 본 아유르바르와다 카안 (원인종(元仁宗), 재위 1311-1320) 시기의 원(元) 조정(朝廷). 중앙아시아연구, 22(2), 73-96.
<https://doi.org/10.29174/cas.2017.22.2.002>
- 조인희 (2020). 정유재란사(丁酉再亂史) 연구에 관한 한일학계 연구의 현황. 학림, 46, 153-204.
<https://doi.org/10.36274/hakrim.2020.46..153>
- 조지형 (2001). 인문학의 '위기'와 디지털 인문학. 김도훈 편. 디지털 시대의 인문학, 무엇을 할 것인가. 서울: 사회평론, 155-196.
- 주성지 (2018). 역사대중화와 디지털 역사자료: 역사소비의 변곡. 역사민속학, 55, 23-54.
<https://doi.org/10.22792/jkhf.2018..55.002>
- 하우봉 (2011). 동아시아 국제전쟁으로서의 임진전쟁. 한일관계사연구, 39, 331-368.
- 하우봉 (2012). 김성일의 일본인식과 귀국보고. 한일관계사연구, 43, 127-174.
- 한국학중앙연구원 (2009). 한국민족문화대백과사전. 한국민족문화대백과사전.
출처: <http://encykorea.aks.ac.kr/>

- 한명기 (2002). 임진왜란기 明·日의 협상에 관한 연구. *국사관논총*, 98, 239-266.
- 한문중 (2009). 壬辰倭亂 직전의 국내정세와 韓日關係. *인문과학연구*, 21, 223-243.
- 함정은, 송민 (2015). 인용정보를 고려한 미발견 공공 지식 추출: Swanson의 ABC 모델 재현 및 확장. *정보관리학회지*, 32(2), 87-103. <https://doi.org/10.3743/KOSIM.2015.32.2.087>
- 허지은 (2004). 정응태의 조선무고사건을 통해본 조·명관계. *사학연구*, 76, 169-206.
- AJ, Berenstein., MP, Magariños., A, Chernomoretz., & F, Agüero F. (2016). A multilayer network approach for guiding drug repositioning in neglected diseases. *Plos Neglected Tropical Diseases*, 10(1), 1-33. <https://doi.org/10.1371/journal.pntd.0004300>
- Barde, B. V. & Bainwad, A. M. (2017). An overview of topic modeling methods and tools. lower dimensional space to find sameness with the docuInternational Conference on Intelligent computing and control systems, Madurai, India. <https://doi.org/10.1109/ICCONS.2017.8250563>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Burdick, Anne., Drucker, Johanna., Lunenfeld, P., Presner, T., & Jeffrey, S. (2012). *Digital humanities*. Massachusetts: MIT Press.
- Ding, Y., Song, M., Han, J., Yu, Q., Yan, E., Lin, L., & Chambers, T. (2013). Entitymetrics: Measuring the impact of entities. *Plos One*, 8(8), e71416. <https://doi.org/10.1371/journal.pone.0071416>
- Eunjeon (2018, July 20). *mecab-ko-dic*. bitbucket. Available: <https://bitbucket.org/eunjeon/mecab-ko-dic/src/master/>
- Lee, M. & Song, M. (2020). Incorporating citation impact into analysis of research trends. *Scientometrics*, 124, 1191-1224.
- Lee, M. (2021, September 6). *tomotopy*. GitHub. Available: <https://bab2min.github.io/tomotopy/v0.12.2/en/>
- Mikko, K., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203-271. <https://doi.org/10.1093/comnet/cnu016>
- Mimno, D. & McCallum, A. (2008). Topic models conditioned on arbitrary features with dirichlet-multinomial regression. *The Twenty-fourth Conference on Uncertainty in Artificial Intelligence*, Helsinki, Finland, 411-418.
- National Institute of Korea History (2021, October 28). *The annals of the joseon dynasty dictionary*. Public data portal. Available: <https://www.data.go.kr/data/15053645/fileData.do>

- Newman, D., Lau, J., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence human language technologies. The 2010 Annual Conference of the North American Chapter of the ACL, Los angeles, USA.
- Song, M. (2021, June 6). treform. GitHub. Available: <https://github.com/MinSong2/treform>
- Steyvers, M. & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, S. D. McNamara, & W. Kintsch eds. Latent semantic analysis: a road to meaning. Laurence Erlbaum, 1-15.
- Vugt, I. v. (2017). Using multi-layered networks to disclose books in the republic of letters. Journal of Historical Network Research, 1(1), 25-51.
- Yu, Q., Ding, Y., Song, M., Song, S. J., Liud, J., & Zhange, B. (2015). Tracing database usage: Detecting main paths in databaselink networks. Journal of Informetrics, 9(1), 1-15.
- 李啓煌 (1997). 文祿.慶長の役と東アジア. 臨川書店.
亂中雜錄.
朝鮮王朝實錄.
懲毖錄.

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Annals of King Seonjo.
- Bang, K. C. (2004). A study on the hakbong Kim Sung Il's conception about Japan. The Journal of the Humanities for Unification, 42, 21-37.
- Cho, I. H. (2020). The history of research on the jungyu war in Korea and Japan. Haklim, 46, 153-204. <https://doi.org/10.36274/hakrim.2020.46..153>
- Cho, J. H. (2001). "Crisis" in humanities and digital humanities. In Kim, Do-hoon (Ed.), Humanities in the digital age. What Will They Do? (pp. 155-196). Seoul: Sapyoungbook.
- Cho, W. H. (2017). New insight on the age of ayurbarwarda (Ch. Renzong, r. 1311-1320) based on a network analysis of officials. Central Asian Studies, 22(2), 73-96. <https://doi.org/10.29174/cas.2017.22.2.002>
- Chung, E. K. (2021). An investigation on digital humanities research trend by analyzing the papers of digital humanities conferences. Journal of the Korean Library and Information Science, 55(1), 393-413. <https://doi.org/10.4275/KSLIS.2021.55.1.393>
- Ha, W. B. (2011). The imjin war as an international war in east asia. The Korea-Japan Historical

- Review, 39, 331-368.
- Ha, W. B. (2012). Kim sungil's japanese recognition and report. *The Korea-Japan Historical Review*, 43, 127-174.
- Ham, J. E. & Song, M. (2015). Detection of hidden knowledge using a citation-based approach based on swanson's abc model. *Journal of the Korean Society for Information Management*, 32(2), 87-103. <https://doi.org/10.3743/KOSIM.2015.32.2.087>
- Han, M. G. (2002). A study on the negotiations between ming and japanese in imjin war. *National Academy's Debate*, 98, 239-266.
- Han, M. J. (2009). Korea's domestic situation and relation with japan prior to the im-jin-wae-ran. *Studies in Humanities*, 21, 223-243.
- Her, J. E. (2004). The choson-ming relations through ding yingtai's false charge against choson. *Sahak Yeongu*, 76, 169-206.
- Hitoshi, N. (2009). A historical review of the study for bunroku keichō no eki. In Korea-Japan Joint History Reach Committee eds, 2nd Korea-Japan Joint History Reach Committee report. Seoul: Korea-Japan Joint History Reach Committee, 237-256.
- Jeong, Y. K. (2020). An analysis on research trends of digital humanities. *Journal of the Korean Society for Information Management*, 37(2), 311-331. <https://doi.org/10.3743/KOSIM.2020.37.2.311>
- Joo, S. J. (2018). Popularizing history and digitized historical materials: A turning point of history consumption. *Journal of Korean Historical-Forklife*, 55, 23-54. <https://doi.org/10.22792/jkhf.2018.55.002>
- Jung, O. K. (2007). Yi hang-bok's diplomatic activities. *The Research for Korean Historical Figures*, 8, 47-72.
- Kang, B. I., Song, M., & Jho, W. S. (2013). A study on opinion mining of newspaper texts based on topic modeling. *Journal of the Korean Society for Library and Information Science*, 47(4), 315-334. <https://doi.org/10.4275/KSLIS.2013.47.4.315>
- Kim, B. R. (2014). Trend of digital humanities. *Humanities Contents*, 33, 229-254.
- Kim, B. R. (2021). Dictionaries and data in the artificial intelligence era exploring for the Korean historical dictionary (data). *The Oriental Studies*, 84, 219-240. <https://doi.org/10.17320/orient.2021.84.219>
- Kim, H. S. (2014). Kim seong-il's activities during the japanese invasion of Korea and his network. *Journal of Nammyung Studies*, 41, 61-98. <https://doi.org/10.14381/NMH.2014.41.03.30.61>

- Kim, H. Y. (2012). The alteration of east asia world order in 16th century and the transition of joseon society around imjin war. *The Korean Journal of History of Historiography*, 26, 87-116.
- Kim, H., Yim, Y. S., & Kim, B. R. (2016). *Introduction to Digital Humanities*. Seoul: HueBooks.
- Kim, K. T. (2014). *A Study on Strengthening Negotiations During the Imjin War*. Doctoral dissertation, Korea University, Korea.
- Kim, K. T. (2019). A study on the secret negotiations between joseon and the japanese army during the imjin war. *The Korea-Japan Historical Review*, 61, 125-172.
<https://doi.org/10.18496/kjhr.2018.08.61.125>
- Kim, M. J. (2017). Forgery and modification of the sovereign's message by tokugawa ieyasu and east asia. *The Korea-Japan Historical Review*, 58, 213-254.
<https://doi.org/10.18496/kjhr.2017.12.58.213>
- Kim, Y. H. & Kim, Y. S. (2019). Trend analysis of healthcare research in Korea using topic modeling. *Journal of Wellness (KSW)*, 14(1), 253-262.
<https://doi.org/10.21097/ksw.2019.02.14.1.253>
- Kwahk, K. Y. (2017). *Social Network Analysis*. Seoul: Crbooks.
- Kwon, Y. K. (2018). Toward a new republic of letters: new attempts of digital humanities in eighteenth-century french history. *Yoksa Hakbo*, 240, 35-74.
<https://doi.org/10.16912/tkhr.2018.12.240.35>
- Lee, G. W. (2009). Eyes of imjin war. In Northeast asian history foundation eds. *Relationship between korea and japan in history*. Seoul: NAHF, 122-143.
- Lee, G. W. (2015). *Early Modern Japan History*. Seoul: Hyeon.
- Lee, J. Y. (2018). Development and recent trend of the digital history in the united states. *Yoksa Hakbo*, 240, 99-126. <https://doi.org/10.16912/tkhr.2018.12.240.99>
- Lee, J. Y. (2018). From humanities in the digital age to the age of digital humanities: history of adopting dh in korean literary studies. *Yoksa Hakbo*, 240, 157-190.
<https://doi.org/10.16912/tkhr.2018.12.240.157>
- Lee, S. H. (2010). Explanatory notes collection of historical documents of the Hideyoshi invasion of Korea(1592-1598). In Korea-Japan Relations History Research and Compilation Committee eds. *East Asian World and the Imjin War*. Seoul: Kyungin, 195-260.
- Lee, Y. S. (2012). The political activities of oem doosoo yoon and japanese invasion of Korea in 1592. *The Research for Korean Historical Figures*, 17, 69-108.
Miscellaneous Record of the War.

- Mun, S. H. (2018). Analytical study on the research trend of the digital history in germany. *Yoksa Hakbo*, 240, 1-33. <https://doi.org/10.16912/tkhr.2018.12.240.1>
- Nam, C. H. (2016). An illustrative application of topic modeling method to a farmer's diary. *Cross-cultural Studies*, 22(1), 89-135.
- Park, E. J. (2018). The development of digital history in britain. *Yoksa Hakbo*, 240, 75-98. <https://doi.org/10.16912/tkhr.2018.12.240.75>
- Park, J. K., Kim, T. K., & Song, M. (2017). Entitymetrics analysis of the research works of dong-ju yun using textmining. *Journal of the Korean Biblia Society for Library and Information Science*, 28(1), 191-207. <https://doi.org/10.14699/kbiblia.2017.28.1.191>
- Roh, Y. K. (2010). Scholarly historical theory review of the hideyoshi invasion of korea. In *Compilation Committee for Research and Discussion on Korea-Japan Relationships eds. East Asian world and Imjin War*. Seoul: Kyungin Publishing, 1-35.
- Song, I. J. (2018). The developments and evolution of digital humanities in taiwan. *Yoksa Hakbo*, 240, 127-155. <https://doi.org/10.16912/tkhr.2018.12.240.127>
- The Academy of Korean Studies (2009). *Encyclopedia of Korean Culture*. Available: <http://encykorea.aks.ac.kr/>
- Writings on Imjin War.
- Yu, J. C. (1986). A study of the korean-japanese diplomatic restoration and the fabrication of national diplpmatic papers from directly after Japanese invasion of Choson(1598) to 1624. *Kangweon Sahak*, 2, 153-354.

[별첨]

〈별첨 표 1〉 형태소 분석기별 성능 비교

형태소 분석기	분석결과
Komorana	[('상', 'NNG'), ('이', 'JKS'), ('평양', 'NNP'), ('을', 'JKO'), ('떠나', 'VV'), ('아', 'EC'), ('영변', 'NNP'), ('부', 'NNB'), ((' ', 'SS'), ('寧邊府', 'SH'), (' ', 'SS'), ('로', 'NNG'), ('향하', 'VV'), ('았', 'EP'), ('다', 'EF'), (' ', 'SF'), ('윤두수', 'NNP'), (' ', 'SP'), ('김명원', 'NNP'), (' ', 'SP'), ('이원익', 'NNP'), ('은', 'JX'), ('머물러', 'VV'), ('어', 'EC'), ('평양', 'NNP'), ('을', 'JKO'), ('지키', 'VV'), ('고', 'EC'), ('대신', 'NNG'), ('최홍원', 'NNP'), (' ', 'SP'), ('유홍', 'NNP'), (' ', 'SP'), ('정철', 'NNP'), ((' ', 'SS'), ('鄭澈', 'SH'), (' ', 'SS'), ('등', 'NNB'), ('은', 'JX'), ('수행', 'NNG'), ('하', 'XSV'), ('았', 'EP'), ('다', 'EC')]
Kkma	[('상', 'NNG'), ('이', 'JKS'), ('평양', 'NNP'), ('을', 'JKO'), ('떨', 'VV'), ('나', 'ECE'), ('영', 'NNG'), ('변', 'NNG'), ('부', 'NNG'), ((' ', 'SS'), ('寧邊府', 'OH'), (' ', 'SS'), ('로', 'NNG'), ('향', 'NNG'), ('하', 'XSV'), ('였', 'EPT'), ('다', 'EFN'), (' ', 'SF'), ('윤두수', 'NNG'), (' ', 'SP'), ('김명원', 'NNG'), (' ', 'SP'), ('이원익', 'NNG'), ('은', 'JX'), ('머물', 'VV'), ('러', 'ECD'), ('평양', 'NNP'), ('을', 'JKO'), ('지키', 'VV'), ('고', 'ECE'), ('대신', 'NNP'), ('최홍원', 'NNG'), (' ', 'SP'), ('유', 'NNG'), ('홍', 'NNG'), (' ', 'SP'), ('정철', 'NNG'), ((' ', 'SS'), ('鄭澈', 'OH'), (' ', 'SS'), ('등', 'NNB'), ('은', 'JX'), ('수행', 'NNG'), ('하', 'XSV'), ('였', 'EPT'), ('다', 'EFN')]
Korean Twitter	[('상이', 'Noun'), ('평양', 'Noun'), ('을', 'Josa'), ('떠나', 'Verb'), ('영변', 'Noun'), ('부', 'Noun'), ((' ', 'Punctuation'), ('寧邊府', 'Foreign'), (' ', 'Punctuation'), ('로', 'Noun'), ('향', 'Noun'), ('하였다', 'Verb'), (' ', 'Punctuation'), ('윤두수', 'Noun'), (' ', 'Punctuation'), ('김명원', 'Noun'), (' ', 'Punctuation'), ('이원익', 'Noun'), ('은', 'Josa'), ('머물러', 'Verb'), ('평양', 'Noun'), ('을', 'Josa'), ('지키고', 'Verb'), ('대신', 'Noun'), ('최홍원', 'Noun'), (' ', 'Punctuation'), ('유홍', 'Noun'), (' ', 'Punctuation'), ('정철', 'Noun'), ((' ', 'Punctuation'), ('鄭澈', 'Foreign'), (' ', 'Punctuation'), ('등', 'Noun'), ('은', 'Josa'), ('수행', 'Noun'), ('했다', 'Verb')]
Mecab	[('상이', 'NNP'), ('평양', 'NNP'), ('을', 'JKO'), ('떠나', 'VV+EC'), ('영', 'NNP'), ('변부', 'NNP'), ((' ', 'SSO'), ('寧', 'SH'), ('邊', 'NNG'), ('府', 'NNG'), (' ', 'SSC'), ('로', 'NNP'), ('향하', 'NNG'), ('였', 'VCP+EP'), ('다', 'EF'), (' ', 'SF'), ('윤두수', 'NNP'), (' ', 'SC'), ('김명원', 'NNP'), (' ', 'SC'), ('이원익', 'NNP'), ('은', 'JX'), ('머물러', 'VV+EC'), ('평양', 'NNP'), ('을', 'JKO'), ('지키', 'VV'), ('고', 'EC'), ('대신', 'NNG'), ('최홍원', 'NNP'), (' ', 'SC'), ('유홍', 'NNP'), (' ', 'SC'), ('정철', 'NNP'), ((' ', 'SSO'), ('鄭', 'NNG'), ('澈', 'SH'), (' ', 'SSC'), ('등', 'NNB'), ('은', 'JX'), ('수행', 'NNG'), ('했', 'XSV+EP'), ('다', 'EC')]

〈별첨 표 2〉 선조실록·수정실록에 대한 중심성 기반 상위 10개 개체

순위	노드	근접 중심성	노드	매개 중심성	노드	고유벡터 중심성
1	판서	0.553108	판서	346346.6	판서	1
2	총병	0.531869	총병	267213.5	유성룡	0.871499
3	유성룡	0.525832	장관	209231.7	총병	0.867269
4	장관	0.523068	유성룡	176189.3	장관	0.827891
5	이덕형	0.51693	경리	174624.3	김응남	0.806102
6	상서	0.513767	이덕형	147837.6	이덕형	0.782408
7	김응남	0.513669	어사	147072.2	상서	0.768629
8	윤두수	0.50948	상서	145227.2	김수	0.750803
9	김수	0.508132	장사	144072.3	윤두수	0.735532
10	장사	0.506981	장령	134477.4	장사	0.707019

〈별첨 표 3〉 징비록에 대한 중심성 기반 상위 10개 개체

순위	노드	근접 중심성	노드	매개 중심성	노드	고유벡터 중심성
1	김명원	0.514286	김명원	2341.325	김명원	1
2	권율	0.418605	권율	884.993	이원익	0.499329
3	심유경	0.416667	왕필적	635.3998	관백	0.454972
4	지사	0.412844	원균	494.0734	이일	0.397529
5	이빈	0.4	심유경	449.0756	평행장	0.380429
6	이일	0.393013	이빈	404.08	윤두수	0.377358
7	이원익	0.391304	지사	385.0714	심유경	0.376288
8	곽재우	0.376569	이일	384.9675	지사	0.373168
9	윤두수	0.371901	유정	325.8848	이빈	0.360344
10	왕필적	0.368852	사대수	263	현소	0.357464

〈별첨 표 4〉 난중잡록에 대한 중심성 기반 상위 10개 개체

순위	노드	근접 중심성	노드	매개 중심성	노드	고유벡터 중심성
1	이광	0.421169	이광	30956.61	이광	1
2	판서	0.383534	판서	18004.45	참장	0.792178
3	곽재우	0.387817	곽재우	15037.92	곽재우	0.748473
4	청정	0.353377	청정	13004.53	정인홍	0.736116
5	참장	0.346328	참장	12631.82	김수	0.705893
6	현령	0.360038	현령	12030.55	김천일	0.662071
7	진이	0.389002	진이	11884.5	고경명	0.661407
8	곽영	0.364504	곽영	11269.82	김성일	0.657225
9	김수	0.383534	김수	11087.83	김면	0.642663
10	서예원	0.38392	서예원	10669.29	서예원	0.617695