

북한의 발해 연구사 서술을 위한 새로운 접근*

— 북한 학술지 목록화 사업 결과물을 활용한 내용분석

고 일 흥**

[초 록]

전통적인 방식의 연구사 서술은 정성분석에 주로 의존하여 주관적인 측면이 강하고, 정량분석도 일면적으로 진행되어 북한 학계의 발해사 연구동향 검토를 위해서는 연구방법의 다각화가 필요하다. 사회과학분야에서 널리 사용되고 있는 다양한 ‘데이터 분석’ 방법론에

* 이 논문은 서울대학교 통일평화연구원 2020년 통일기반구축사업의 지원을 받아 수행되었다.

** 서울대학교 아시아연구소 HK연구교수

이 글의 토대가 되는 북한의 발해 연구 분석을 위한 기초 데이터는 북한 인문학 시맨틱 데이터 아카이브 연구 웹사이트(<http://dh.aks.ac.kr/~nkh/index.html>) 상단의 탭 가운데 Data 항목 하위의 Dataset을 통해 다운로드하여, 본 연구의 내용을 구체적으로 검증할 수 있음을 밝힌다. 또한 본 논문의 인쇄본에는 원래 컬러로 제작된 도면들이 흑백으로 출력되어 필자의 의도가 제대로 전달되지 않아, 이 웹사이트를 통해 본 논문의 컬러 도면도 제공되고 있다.

주제어: 북한 학술지, 발해사, 내용분석, 디지털 데이터 분석, 시맨틱 네트워크 분석, 데이터 시각화

North Korean Research Journals, Balhae, Content Analysis, Digital Data Analysis, Semantic Network Analysis, Data Visualization

서 새로운 돌파구를 찾아, 본고에서는 북한 학술지 목록화 사업의 일환으로 집적된 자료에 대한 정량적/개념적 내용분석 및 정성적/관계적 내용분석을 진행하여 북한의 발해사 연구동향을 파악하기 위한 새로운 접근방법을 제시하였다. 이를 위해 ‘다량의 균질적 데이터’의 성격을 가진 북한 학술지의 논문 제목을 분석대상으로 삼았고, 이에 대한 통계적 분석과 초보적인 수준의 시맨틱 네트워크 분석(Semantic Network Analysis)을 진행하였다. 우선, 정량적/개념적 내용분석의 일환으로 북한 발해사 연구 논문의 발간 현황을 통시적으로 살펴보고, 『조선고고연구』와 『력사과학』으로 대변되는 고고학계와 역사학계의 발해사 연구동향을 비교분석하고, 고구려사 및 고조선사 논문의 발간 현황과의 비교분석도 진행하였다. 그 결과, 2000년대부터 발해 논문을 수록한 학술지의 다변화가 일어났고, 발해사 연구의 주요 주제로 여겨져 왔던 ‘계승’ 키워드의 논문들이 1997~98년도에 집중되어 있고, 그 이후 발해 연구 담론을 주도하는 역할을 했다고 볼 수 없다는 사실이 확인되었다. 정성적/관계적 내용분석을 위해서는 넷마이너 4.0 소프트웨어를 이용하여 발해 논문 제목의 개념어들이 ‘노드’로 설정된 ‘시맨틱 네트워크’를 생성하고, 그 결과를 ‘내용 클라우드’(content cloud)의 형태로 시각화하였다. 기존보다 훨씬 더 많은 데이터를 한 눈에 살펴볼 수 있게 해주는 이러한 내용 클라우드를 통해 북한 발해사 연구의 관련 개념어들을 확인하였고, 학술지별 비교도 진행하였다. 또한, 시맨틱 네트워크에 대한 연결중심성(degree centrality) 분석을 통해 발해사 연구의 중심 키워드들을 확인하였고, 역시 학술지별 비교도 진행하였다. 결국 이러한 연구를 통해 북한 학계의 발해사 연구경향에 대한 이해의 폭을 확대하고, 목록화 사업으로 집적된 자료의 활용 방안을 모색하고, 인문학적 연구에 디지털 데이터 분석 기술을 적용한 사례를 제시할 수 있었다.

1. 머리말

해방 이후부터 1970년대까지 발해사 연구는 북한 학계에 의해 주도되었다.¹⁾ 발해가 모든 측면에서 고구려를 계승한 것으로 보는 발해사 연구의 기본 관점은 1962년에 『력사과학』을 통해 제시되었고, 이후 북한 연구자들의 성과를 기반으로 발해사와 한국사가 연결될 수 있도록 하는 틀이 마련되었다.²⁾ 또한 북한의 함경남북도 지역은 발해사 연구의 기본이 되는 물질자료(즉, 유적과 유물)의 핵심 발견 지역 중 하나이다.³⁾ 따라서 남한의 발해사 연구자들은 북한 학계의 연구동향에 지대한 관심을 가졌으며,⁴⁾ 심지어 『신편 한국사』에서도 북한의 발해사 인식과 연구가 언급되었다.⁵⁾

-
- 1) 한규철(1994), 「북한의 발해사 연구」, 『北韓의 古代史研究와 성과』, 서울: 대륙연구소, p. 120.
 - 2) 송기호(2012), 「발해의 문화와 발해사 인식의 변천」, 『발해(신편 한국사 10)』, 서울: 국사편찬위원회, p. 238.
 - 3) 이동휘(2020), 「북한 경내의 발해유적 발굴조사성과와 그 의의 — 부거리와 북청 일대 —」, 『역사와 세계』 57(1), 효원사학회.
 - 4) 2001년 이전까지의 남한에서 나온 북한의 발해사 연구 논고들은 국사편찬위원회(2001), 『북한 역사학 논저 목록(하)』, 서울: 국사편찬위원회, pp. 691-694에 일부 정리되어 있다. 이 외에도 다음과 같은 논고들이 있다: 송기호(1990), 「북한의 발해 고고학과 《발해문화》」, 『역사와 현실』 3, 한국역사연구회; 송기호(1991), 「北韓의 渤海史·統一新羅史 研究」, 『北韓의 古代史研究』, 서울: 一潮閣; 이인철(2003), 「북한의 후기신라·발해사 연구동향」, 『북한의 한국사 연구동향(1)』, 서울: 국사편찬위원회; 이병건(2007), 「발해건축에 관한 북한의 연구성과」, 『高句麗渤海研究』 26, 고구려발해학회; 이병건(2018), 「북한의 발해건축 관련 연구와 향후 우리의 과제」, 『高句麗渤海研究』 61, 고구려발해학회. 한편, 북한 학계의 고구려 및 발해사의 최신 연구성과를 소개하는 노력들도 이어지고 있다. 예를 들어, 정경일·王寧(2019), 「2019년 북한학계의 고구려 및 발해사 연구 목록」, 『高句麗渤海研究』 65, 고구려발해학회; 정경일·김명(2019), 「2019년 북한학계의 고구려 및 발해사 연구 목록」, 『高句麗渤海研究』 65, 고구려발해학회.
 - 5) 송기호(2012).

북한의 발해사 연구경향을 다룬 남한 학계의 기존 성과들은 대체로 정성분석에 의존하였다.⁶⁾ 즉, 대표 연구자 및 연구결과를 중심으로 연구의 큰 흐름을 설명하고, 그 의의와 한계에 대한 평가가 내려졌다. 그런데 이러한 전통적인 방식의 ‘연구사 서술’에서는 모든 자료에 대한 정성분석과 기술이 불가능하다 보니 ‘대표사례’를 중심으로 서술이 이루어지고, 주류에서 벗어난 경향들은 간과될 가능성이 높다. 또한, 연구사를 쓰는 작업은 단순히 기술(記述)에만 머물지 않고, 학계 담론의 진행방향에 영향을 주고자 하는 의도를 내포할 수 있다.⁷⁾ 그런데 전통적인 방식으로 서술된 연구사에서는 참고문헌 외에는 검증의 수단이 제시되고 있지 않아, 해당 분야에 대한 지식이 충분하지 않은 독자는 집필자가 제시하는 객관적 사실과 집필자의 정치적 의도를 담은 주관적 해석을 구분하지 못할 수도 있다.

위와 같은 문제점을 보완하기 위해 ‘데이터 분석’ 방법론을 활용하여 북한의 발해사 연구경향을 검토하는 것이 본고의 목표이다. 이를 위해 기술통계와 같은 계량적 접근방법을 적용하고, 비록 초보적인 수준이지만 시맨틱 네트워크 분석(Semantic Network Analysis)을 시도하고자 한다. 분석 대상이 되는 ‘데이터’는 ‘통일기반구축사업’의 일환으로 목록화가 진행된 북한 고고학 및 역사학의 대표 학술지인 『력사과학』, 『조선고고연구』, 『민족문화유산』 등에 수록된 논문의 제목들이다. ‘제목’에 초점을 맞춘 이유는 학술논문에서 제목이 흔히 1) 논문의 내용에 대한 명료한 메시지 전달, 2) 독자의 관심 유도, 3) 주제어를 별도로 제시하지 않는 경우에는 주제어 제시⁸⁾와 같은 기능을 수행하는 만큼,

6) 예외적으로 계량분석이 부분적으로나마 진행된 성과로 이인철(2003), 앞의 글의 pp. 115-116쪽 내용을 들 수 있다.

7) Graham, L., W. Lепенies, & P. Weingart (eds.)(1983), *Functions and Uses of Disciplinary Histories*, Dordrecht: D. Reidel Publishing Company.

8) Goodman, R. A., S. B. Thacker, & P. Z. Siegel (2001), “What’s in a Title? A Descriptive Study of Article Titles in Peer-Reviewed Medical Journals”, *Science*

집필자가 의도한 논문의 내용과 성격을 가장 잘 전달해주는 데이터로 볼 수 있기 때문이다. 또한 논문의 본문까지 분석하는 심도 있는 접근 대신에 제목에 초점을 맞춘 분석을 진행하면 좀 더 광범위한 시야를 확보하여 북한의 발해사 연구경향을 고구려사 및 고조선사 연구경향과 비교하는 연구도 가능해진다.

이에 본고의 제2장에서는 분석 대상 자료의 선정 근거를 제시하고, 활용하고자 하는 분석 방법론의 가능성에 대해 설명하고자 한다. 제3장에서는 계량적 및 정성적 내용분석의 과정과 결과를 제시하고, 제4장에서는 이러한 새로운 접근의 결과물이 어떻게 발해사 연구경향에 관한 기존의 서술과 상호보완적인 관계를 맺을 수 있는지를 살펴보도록 하겠다. 이와 같은 시도를 통해 북한 학계의 발해사 연구경향에 대한 이해의 폭을 넓히는 것은 물론, 그간 진행된 다양한 목록화 사업을 통해 축적된 자료의 새로운 활용 방안을 찾고, 또한 디지털 기술을 인문학적 연구에 접목시켜 디지털 인문학의 저변 확대에 조금이나마 기여하는 것이 필자가 본고를 통해 궁극적으로 추구하는 바이다.

2. 북한의 학술지 데이터 분석을 위한 새로운 틀

2.1. 분석 대상

북한의 발해사 연구경향을 다룬 기존의 연구들은 저서와 논문을 모두 분석의 대상으로 삼았다. 이는 저서와 논문에서 다루어지는 연구 범위, 심도, 주제 등은 서로 다른 만큼, 두 종류의 자료를 같이 보아야만 전반적인 연구경향에 대한 정확한 파악과 해석이 가능해지기 때문

이다. 본고의 경우에는 기존 시도들과의 차별화를 위해 ‘데이터 분석’을 통한 발해사 연구경향의 통시적 변화양상 검토, 발해사·고구려사·고조선사 연구경향 비교검토, 고고학계·역사학계 주도 발해사 연구경향의 비교검토를 진행하고자 한다. 그런데 이러한 분석을 위해서는 다량의 균질적인 데이터가 필요하다.

연구결과의 출판과 공유를 위한 가장 중요한 공적인 수단에 해당되는 학술지는 그 특성상 주기적으로 생산된 유사한 분량의 다양하고 많은 연구를 포함하고 있으며, 통시성도 담보하고 있다. 따라서 ‘다량의 균질적인 데이터’가 요구되는 본고의 새로운 접근을 위해서는 이러한 학술지에 수록된 논문이 적합한 연구의 대상인 것으로 판단되었다.

북한에서 발행된 발해 논문 수록 대표 학술지로는 『조선고고연구』, 『력사과학』, 『민족문화유산』, 『고고민속』, 『조선건축』, 『김일성종합대학학보(역사·법률)』이 있으며, 각 학술지의 발행 정보는 <표 1>에 제시되어 있다. 북한의 발해사 연구 그 자체에 대한 이해의 확보를 위해 본고에서는 우선 이 모든 학술지에 수록된 발해 논문을 분석의 대상으로 삼도록 하겠다. 또한 북한 고고학·역사학계의 연구경향 속에서 발해 관련 연구에 대한 맥락화를 위해 이 중에 10년 이상 지속적으로 발행된 고고학 및 역사학 분야의 『조선고고연구』와 『력사과학』에 수록된 모든 — 즉, 발해 및 非발해 — 논문도 분석의 대상으로 삼도록 하겠다.

〈표 1〉 본고에서 분석대상으로 삼은 발해 논문 수록 학술지 정보

학술지 이름	발행기간	발행주기	발행기관
『조선고고 연구』	1986~2019	계간	·사회과학원 고고학연구소
『역사과학』	1955~2019	월간(1955년) 격월간(1956~1967) 계간(1977~2018) ⁹⁾	·1955~1988년 3월 이전: 과학백과사전출판사 ·1988년 3월 이후: 사회과학출판사
『민족(문화) 유산』 ¹⁰⁾	2001~2019	계간	·2001~2003년 제4호: 조선문화보존사 ·2004년 제1호 이후: 과학백과사전출판사공업출판사
『고고민속』	1963~1967	계간	·1963년 1호~1964년 1호: 조선민주주의인민공화국과학원 ·1964년 2호~1966년 4호: 사회과학원 고고학 및 민속학연구소 ·1967년 간행물: 사회과학원 고고학연구소
『조선건축』	1987~2019	계간(1997년 4호~2003년까지 발행 중지 기간으로 간행되지 않음)	·공업출판사
『김일성 종합대학학보: 력사학』	1993~2019	계간(연5회 간행되는 경우도 있음)	·1993~2010년: ‘력사, 법학’ 부제 ·2011~2018년 제3호: ‘력사, 법률’ 부제 ·2018년 제4호: ‘력사학’ 부제

9) 북한의 발해사 연구를 다룬 다수의 남한 문헌에서는 『역사과학』 1976년 3호에 실린 박시형의 논고 「발해력사는 조선력사의 한 부분이다」를 언급하고 있다. 그러나 필자는 『역사과학』 1976년 3호는 물론, 1호나 2호의 목차를 확보할 수 없었다. 통일부 북한자료센터에도 소장되어 있지 않으며, 국립문화재연구소에서 발간한 『북한 정기간행물 — 고고학 문헌 목록』에도 중단 기간 이후 『역사과학』 1977년 1호의 목차부터 제시되어 있다. 따라서 ‘발해 논문 수록 현황’의 파악 등을 위해 박시형의 1976년 논고를 분석대상에 포함시켰으나, 『역사과학』의 발행기간은 이와 같음을 밝혀둔다.

10) 『민족문화유산』에서 2019년 1호 루계 제73호부터 『민족유산』으로 제목이 변경됨.

2.2. 분석 방법

2.2.1. 계량적/개념적 내용분석

‘내용분석’(Content Analysis)이란, 전통적인 정의에 따르면 “의사소통에서 드러난 내용에 대한 객관적, 체계적, 그리고 정량적인 기술(記述)을 위한 연구 방법”이며,¹¹⁾ “메시지의 명시된 특징을 객관적이고 체계적으로 확인함으로써 유추를 하는 모든 방법론”¹²⁾을 포괄한다. 내용분석은 20세기 중반 이후 사회과학·의학·간호학 분야 등에서 널리 사용되었으며, 21세기에 들어와서는 학술지 수록 논문에 대한 내용분석을 통한 연구경향 파악이 이들 분야를 중심으로 본격적으로 이루어졌다.¹³⁾ 인문학 분야의 경우에도 연구경향 파악을 위한 학술지 대상 내용분석이 이루어진 사례들이 있다. 최신 사례로는 A&HCI 학술지에 게재된 ‘한국’ 관련 논문들에 대한 내용분석을 통해 1990~2005년 동안의 한국학 연구경향을 파악한 슈웨이겐디그(Schwekendiek)의 연구¹⁴⁾를 들 수 있다.

내용분석에는 ‘계량적 내용분석’(QnCA: Quantitative Content Analysis)과 ‘정성적 내용분석’(QCA: Qualitative Content Analysis)이 있으며, 전

11) Berelson, B. (1952), *Content Analysis in Communication Research*, Glencoe, Ill.: Free Press., p. 18.

12) Holsti, O. R. (1969), “Content Analysis”, *Content Analysis for the Social Sciences and Humanities*, Boston: Addison-Wesley, p. 14.

13) 예를 들어, 고고학과 인접한 학문인 지리학의 경우에는 다음과 같은 사례가 있다. Jackson, A., R. Harris, L. Hepple, A. Hoare, R. Johnston, K. Jones, & P. Plummer (2006), “Geography’s changing lexicon: measuring disciplinary change in Anglophone human geography through journal content analysis”, *Geoforum* 37.

14) Schwekendiek, D. J. (2020), “Trends in Korean studies: A content analysis of Korea-related articles published in the Arts & Humanities Citation Index, 1990-2015”, *International Area Studies Review* 온라인 게재 (<https://journals.sagepub.com/doi/abs/10.1177/2233865920934954>).

자는 흔히 언론정보학, 경영학, 행정학, 도서관학 분야에서, 후자는 간호학, 교육학, 의학, 심리학 분야에서 널리 사용되어 왔다.¹⁵⁾ 계량적 내용분석은 연구자가 설정한 범주의 해당 관찰내용을 통계 데이터로 전환하는 작업을 동반하는데, 통계 결과를 해석할 때의 기본적인 전제는 불용어(stop words)를 제외한 가장 높은 빈도로 나타나는 단어와 구절이 의사소통 과정에서 가장 중요도가 높은 사항을 반영하고 있다는 것이다.¹⁶⁾

개념적 분석(conceptual analysis) 혹은 주제 분석(thematic analysis)과 일맥상통하는 이 분석방법에서는 연구 관심사와 관련이 있는 ‘개념’을 설정하고, 빈도분석의 대상이 될 관련 ‘개념어’를 규정하는 것이 핵심적인 작업이 되겠다. 따라서 본고의 목적을 위해서는 세 가지 방식으로 계량적/개념적 내용분석을 진행하겠다. 첫째, <표 1>의 모든 학술지에 수록된 ‘발해’ 개념어 포함 ‘글’에 대한 빈도분석을 진행하도록 하겠다. 둘째, 『조선고고연구』와 『역사과학』에 수록된 모든 ‘논문’을 분석의 대상으로 삼고, ‘발해’, ‘고구려’, ‘고조선’을 개념어로 규정하여 빈도분석을 진행하도록 하겠다. 셋째, 발해 연구의 대표 ‘주제 범주’들을 개념어를 규정하고, 『조선고고연구』와 『역사과학』에 수록된 발해 논문들을 대상으로 그 개념어들의 빈도분석을 진행하도록 하겠다.

2.2.2. 정성적/관계적 내용분석

정성적 내용분석은 “체계적인 범주화 과정인 코딩과 주제 혹은 패턴의 확인을 통해 텍스트 데이터에 대한 주관적 해석을 진행하는 방법론”¹⁷⁾으로 정의될 수 있으며, 이러한 작업은 “일련의 정성적 자료를

15) Prasad, B. D. (2019), “Qualitative Content Analysis: Why is it Still a Path Less Taken?”, *Forum: Qualitative Social Research* 20(3).

16) Prasad (2019).

17) Hsieh, H.-F. & S. E. Shannon (2005), “Three approaches to qualitative content analysis”,

대상으로 핵심적인 일관성과 의미를 확인하기 위해 진행되는 데이터 간략화(data reducing)의 노력 및 그 결과를 해석하기 위한 노력으로 볼 수 있다.¹⁸⁾ 따라서 정성적 내용분석은 결과물은 그래프로 나타내고 통계를 처리할 수 있는 숫자 데이터가 아니라, 텍스트로부터 귀납적으로 얻어진 명목적 데이터이다.¹⁹⁾

관계 분석(relational analysis) 혹은 의미 분석(semantic analysis)과 일맥상통하는 이 분석방법 역시 연구 관심사와 관련이 있는 ‘개념’의 설정부터 시작한다. 하지만 그러한 개념어의 존재에 초점을 맞추기 보다는, 개념어 간의 관계 파악 및 의미 도출이 연구의 지향점이 된다. 이러한 개념어 간의 관계 파악을 위해서는 다양한 소프트웨어 프로그램이 이용된다. 본고의 목적을 위해서는 넷마이너(NetMiner) 4.0 소프트웨어를 활용해서 초보적으로나마 시맨틱 네트워크 분석을 적용하여 정량적/관계적 내용분석을 진행하도록 하겠다. 시맨틱 네트워크 분석(Semantic Network Analysis)이란 텍스트로부터 ‘개념’인 노드를 추출하고 ‘개념 간 인접성 관계’를 나타내는 링크를 파악하여 하나의 ‘시맨틱 네트워크’를 구성함으로써 그 텍스트에 대한 내용분석을 진행하는 방법을 의미한다. 기존의 네트워크 분석에서는 텍스트 데이터를 사전에 노드와 링크로 구조화시킬 필요가 있었으나, 넷마이너 4.0 소프트웨어를 이용해서는 ‘비정형 텍스트 데이터’로부터 노드가 자동으로 추출되고, 이 노드들로 구성된 네트워크 데이터를 생성할 수 있게 되었다.²⁰⁾

이러한 분석을 통해 첫째, 텍스트 내에서 단어가 얼마나 중요한 위치를 차지하는지의 정도를 나타내는 통계적 지표인 TF-IDF를 바탕으

Qualitative Health Research 15(9), p. 1278.

18) Patton, M. Q. (2002), *Qualitative Research and Evaluation Methods*, Thousand Oaks, CA: Sage, p. 453.

19) Prasad (2019).

20) Cyram (2017) 『NetMiner 4.0 Semantic Network Analysis 매뉴얼』, p. 8.

로 텍스트 내 단어들의 상대적 중요성을 시각화한 ‘내용 클라우드’(content cloud)를²¹⁾ 생성하여 발해 관련 연구의 주된 주제들을 파악하도록 하겠다. 정치적 연설 분석에 널리 사용된 바 있는 이 방법은 개념어의 폰트 크기를 통해 그 상대적 중요성을 한 눈에 알아보기 쉽게 표현해 주는데, 정성적 내용분석을 위한 하나의 방법으로 그 가능성이 타진된 바 있다.²²⁾ 둘째, 자동으로 추출된 단어들 간의 관계를 네트워크 맵으로 시각화하고 연결중심성 분석을 진행하여 가장 영향력 있는 발해사 연구의 주제들과 그 관계를 확인하도록 하겠다. 논문 주제어에 대한 네트워크 분석을 통하여 연구경향을 밝히는 시도들이 오늘날 활발히 이루어지고 있으며, 한국 연구진에 의해 진행된 사례도 있다.²³⁾

3. 북한의 학술지 데이터 분석의 새로운 가능성

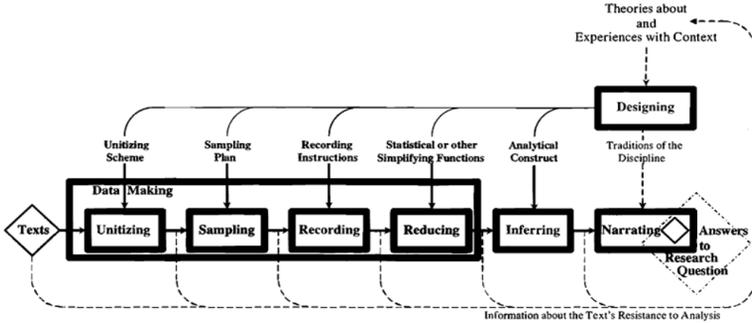
크리펜도르프(Krippendorff)에 의하면 내용분석의 과정은 다음과 같은 단계들로 이루어져 있다. 1) 분석의 틀 디자인하기, 2) 분석의 단위 규정하기, 3) 샘플링하기, 4) 기록/코딩하기, 5) 간략화하기, 6) 추론하기, 7) 서술하기.²⁴⁾ 이러한 과정을 도식화 한 것이 [그림 1]이고, 각 단계의 핵심 고려 사항을 정리한 것이 <표 2>이다. 본 장에서는 이러한 틀에 맞추어 북한의 학술지 데이터에 대한 내용분석의 과정과 결과를

21) Cyram (2017), pp. 12-13.

22) Cidell, J. (2010), “Content clouds as exploratory qualitative data analysis”, *Area* 42(4).

23) Choi, S. H., H. J. Seo, & Y. S. Kim (2016), “Analysis of the Research Trends of the Korean Journal of Educational Research Using Network Text Analysis”, *International Journal of Software Engineering and Its Applications* 10(12).

24) Krippendorff, K. (2004), *Content Analysis: An Introduction to Its Methodology* (2nd ed.), Thousand Oaks, CA.: Sage.



[그림 1] 내용분석의 진행과정.²⁵⁾

<표 2> 내용분석의 단계별 고려 사항²⁶⁾

내용분석의 단계	단계별 고려 사항
분석의 틀 디자인	<ul style="list-style-type: none"> 기존의 논의와 연구자의 경험을 바탕으로 가설을 설정함. 내용분석의 결과가 가설 검증에 유효하도록 함.
분석단위 규정	<ul style="list-style-type: none"> ‘단위’는 분석자가 독립적으로 존재로 요소로 인식하는 대상이며, 단위설정은 분석대상의 개별 요소를 규정하는 행위임. 텍스트의 경우, 장, 문단, 문장, 구절 등이 분석단위가 될 수 있음.
샘플링	<ul style="list-style-type: none"> 분석대상을 연구자가 다룰 수 있는 규모로 줄이되, 그것이 전체를 충실히 반영하도록 하는 전략을 구사해야 함.
기록/코딩	<ul style="list-style-type: none"> 분석대상에서 다양한 방식으로 패턴을 찾아낼 수 있도록 해 줌. 다른 연구자에 의해서도 동일한 패턴의 도출이 가능해야 함.
간략화	<ul style="list-style-type: none"> 분석결과를 쉽게 인지할 수 있도록 정리하는 행위로, 도표 등이 이용됨.
추론	<ul style="list-style-type: none"> 도출된 결과를 바탕으로 분석대상 특정 변수의 향후 추세나 다른 분석대상과의 비교분석을 통한 더 큰 현상을 예측함.
서술	<ul style="list-style-type: none"> 연구자가 속한 학문의 서술전통이나 담론관습에 따라 내용분석을 활용한 연구의 서술방식이 정해져야 함.

25) Krippendorff (2004), p. 86, Figure 4.2 인용.

26) 이 표의 내용은 Krippendorff (2004)의 4~9장의 내용을 정리한 Joram Binsbergen의 ‘Quantitative Content Analysis’ 기사(23, Sep. 2013)를 참고하였음(<https://www.digitalmethods.net/MoM/QuantContentAnalysis>).

제시하도록 하겠다.

3.1. 북한의 학술지 수록 논문 대한 계량적/개념적 내용분석

3.1.1. 분석의 과정

대한민국 학계에서 진행된 북한의 발해사 연구경향 관련 논의를 보면 ‘단계설정’이 흔히 이루어진다. 대표적인 사례를 들면, 북한의 발해 건축 관련 연구를 다룬 최근의 시도에서는 1960년대에서 2000년대 이후까지를 10년 단위로 ‘조사착수기’, ‘연구개척기’, ‘조사확장기’, ‘연구확대기’, ‘조사침체기·연구융합기’로 설정한 바 있다.²⁷⁾ 또한 이에 앞서 한규철은 발해사 연구가 박시형, 주영현에 의해 시작된 이후로, 몇 년간의 공백이 있는 이후 1986년도부터 다시 본격화되었다가, 1990년부터 발해사가 중요한 위치를 차지하게 되었다고 보았으며,²⁸⁾ 송기호는 80년대 후반에 북한의 발해사 연구 기구와 인력이 대폭 확충되면서 연구 성과가 늘어난 것으로 보았다.²⁹⁾ 한편, 이러한 연구들에서는 ‘역사학계’와 ‘고고학계’를 구분하거나,³⁰⁾ 1980년대부터 진행된 주체사상의 강화와 관련지어 북한의 발해사 연구 경향을 서술하고 있다.³¹⁾

따라서 본 연구에서는 첫째, 북한 발해사 연구 논문의 발간 현황을 통시적으로 살펴봄으로써 기존의 논의에서 규정된 ‘공백’, ‘확충’, ‘침

27) 이병건(2018).

28) 한규철(1994), p. 126.

29) 송기호(2012), p. 239.

30) 예를 들어, “1980년대 전반까지의 발해사 연구는 박시형의 무대였다고 할 수 있다. 문헌에서는 손영중의 것(1980)을 제외하고는 모두가 박시형이 쓴 것이었기 때문이다. 그러나 고고학에서는 주영현과 박진옥, 장상렬이 함께 하던 시기이다 … 이미 고고학에서 발해사에 관심을 갖기 시작한 것은 도유호로부터이다.” 한규철(1994), p. 124.

31) 송기호(1990), p. 276.

체'의 양상을 검증하고자 했다. 이를 위해서는 『조선고고연구』, 『역사과학』, 『민족(문화)유산』, 『고고민속』, 『조선건축』, 『김일성종합대학학보(역사·법률)』에 수록된 모든 '발해' 키워드가 등장하는 '논문 제목'을 분석의 대상으로 삼았다.³²⁾

한편, 본 연구에서 '발해'가 등장하는 제목만을 연구대상으로 설정하면서 발해 관련 글이 누락된 경우도 있었다. "청해토성 및 교성리토성 답사보고"³³⁾나 "새로 발굴된 24개돌"³⁴⁾ 등이 여기에 해당된다. 이러한 사례들을 추가하는 방식으로 분석대상 샘플을 보정하지 않은 이유는 논문 제목에 '발해'가 포함되어 있지 않다는 사실이 집필자의 의도나 집필 당시의 학문적 맥락을 나타낸다고 판단했기 때문이다. 따라서 분석대상에는 포함시키지 않았지만, 필요에 따라 분석결과의 추론 혹은 서술 단계에서는 참고하도록 하겠다. 이러한 기준에 따라 선정된 샘플은 여섯 종류의 학술지에 수록된 총 191편의 '글' 제목이다(표 3 참고).

본 연구에서는 둘째, 발해사 연구의 정도가 주체사상의 강화와 상관관계가 있는지를 살펴보기 위해 역시 주체사상의 영향으로부터 자유로울 수 없는 고구려사 및 고조선사 논문 발간양상과의 비교분석을 진행하였다. 비교분석을 위해 『조선고고연구』와 『역사과학』에 수록된 '발해', '고구려', '고조선' 제목 논문들의 발간 비율을 통시적으로 살펴보았다. 주목할 사항은, 『역사과학』에는 '수령사' 관련 논문이 그해

32) '논문 제목'이 과연 내용분석의 타당한 대상이 될 수 있는지와 관련해서는, 앞서 (p. 2) 설명했듯이, 별도의 주제가 제시되지 않는 경우에는 제목이 주제의 역할을 수행 할 수 있다. 실제로 논문의 제목 대한 내용분석이 진행된 경우가 많으며, 다음과 같은 사례를 참고할 수 있다: Jamali, H. K. & Nikzad, M. (2011), "Article title type and its relation with the number of downloads and citations", *Scientometrics* 88(2); Hudson, J. (2016), "Analysis of the titles of papers submitted to the UK REF in 2014: authors, disciplines, and stylistic details", *Scientometrics* 109.

33) 이정기(1967), 「청해토성 및 교성리토성 답사보고」, 『고고민속』 1967-4, 사회과학원 고고학연구소.

34) 리준걸(1984), 「새로 발굴된 24개돌」, 『역사과학』 1984-2, 과학백과사전출판사.

발간된 논문 편수의 30~50%를 차지하는 경우도 있어서, ‘발해’, ‘고구려’, ‘고조선’ 제목 논문의 비율을 구할 때에는 이러한 ‘수령사’ 관련 논문은 배제하였다. 『조선고고연구』의 경우에는 ‘논문’ 외에도 ‘발굴보고’, ‘강좌’ 등이 수록되어 있는데, 비교분석을 위해서는 오로지 ‘논문’ 범주에 한정하여 ‘발해’(n=58), ‘고구려’, ‘고조선’ 제목 논문의 비율을 구했다³⁵⁾. 또한 『조선고고연구』와 『역사과학』의 통시적 비교를 위해 전자가 창간된 연도인 1986년 이후에 발간된 『역사과학』 논문을 대상으로 분석을 진행하였다.

마지막으로, 본 연구에서는 고고학계와 역사학계의 발해 연구의 방향성에서 어떠한 차이가 존재하는지를 확인하고자 했다. 이를 위해 『조선고고연구』와 『역사과학』에 수록된 ‘발해’ 제목 논문들을 총 4개의 공통된 주제 범주(‘계승’, ‘정치/외교/군사/주민구성’, ‘유적’, ‘유물’)와 각 정기간행물의 특성에 맞는 3개의 추가적인 주제 범주(『조선고고연구』는 ‘주체’, ‘도시/건축/성곽’, ‘무덤’, 『역사과학』은 ‘인식/사료’, ‘영역/위치비정’, ‘경제/문화/도로’)를 설정하여, 발해사 연구 주제의 분포 양상을 통시적으로 살펴보고, 또한 고고학계와 역사학계를 대변하는 각 정기간행물 수록 논문들의 연구 주제의 구성을 비교하였다.

3.1.2. 분석 결과

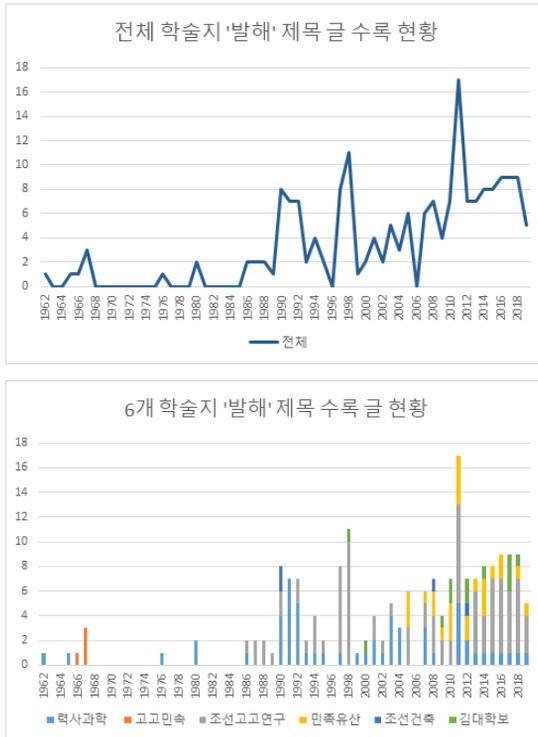
3.1.2.1. 북한 발해사 연구 논문의 발간 추이

북한의 발해사 연구 논문의 발간 추이를 보면 『조선고고연구』에 가장 많은 수의 글(‘논문’, ‘발굴보고’, ‘강좌’, ‘자료’ 포함)이 수록되었고, 『역사과학』과 『민족(문화)유산』이 그 뒤를 잇고 있는데, 이 중에서 『역사과학』의 발행기간이 가장 긴 점을 감안하면 고고학계가 북한의 발

35) 앞서 언급한 총 6종류의 학술지에 수록된 총 192편의 논문을 취합할 때에는 비교 분석이 목적이 아니었기 때문에 『조선고고연구』에서 ‘발굴보고’, ‘강좌’, ‘자료’에 해당되는 글들도 포함시켰다(n=90).

〈표 3〉 북한의 대표 학술지에 수록된, 제목에 ‘발해’가 포함된 모든 ‘글’의 수록 편수

학술지 이름	‘발해’ 제목 ‘글’ 수록 편수
『조선고고연구』	90
『역사과학』	56
『민족(문화)유산』	25
『고고민속』	4
『조선건축』	4
『김일성종합대학학보(역사학)』	12
합 계	191



[그림 2] 〈표 3〉의 모든 학술지(위) 및 학술지별(아래) ‘발해’ 제목 글의 편수와 그 통시적 변화 양상.

해사 연구를 주도하고 있음을 알 수 있다(표 3 참고).

한편, 글의 수록 편수를 통시적으로 살펴보면, 세 개의 피크(peak)가 뚜렷이 확인되는데, 첫 번째인 피크(peak)에 해당되는 1990~1992년 무렵에 역사학계(『역사과학』)를 중심으로, 그리고 두 번째 피크(peak)에 해당되는 1997~1998년 무렵에는 고고학계(『조선고고연구』)를 중심으로 연구가 주도되었음을 확인할 수 있다. 그리고 그 뒤를 이은 2011년의 세 번째 피크(peak)와 그 이후에 매년 5편 전후의 발해 관련 등이 꾸준히 출간된 양상은 그러한 글을 수록한 학술지의 다변화와 관련이 있어 보인다(그림 2 참고).

3.1.2.2. 『조선고고연구』와 『역사과학』의 발해사, 고구려사, 고조선사 논문 비율 비교

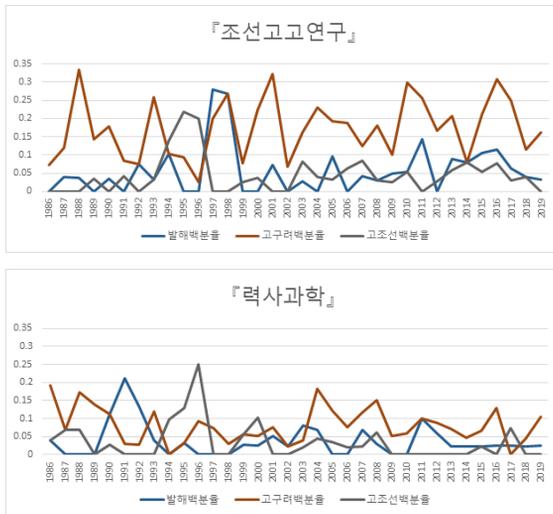
1986년 이후부터 『조선고고연구』와 『역사과학』에 수록된 발해사, 고구려사, 고조선사 논문의 비율을 보면, 두 학술지 모두에서 고구려사 관련 논문이 압도적으로 많으며, 발해사 논문의 비율이 두 번째로 높고, 고조선사 논문의 비율이 가장 낮음을 알 수 있다. 또한 『조선고고연구』의 경우에 고구려사에 대한 편중 양상이 더욱 심하게 확인된다(표 4 참고).³⁶⁾

그런데 흥미롭게도 발해사, 고구려사, 고조선사 논문의 수록 편수를 통시적으로 살펴보면 발해사와 고조선사에 대한 관심이 고구려사에 대한 관심보다 높았던 시기들이 확인된다. 우선, 두 학술지 모두에서 1994~96년도에는 ‘고조선’ 논문의 비율이 ‘고구려’ 논문의 비율보다 높았다. 또한, ‘발해’ 논문의 경우에는, 『조선고고연구』에서는 1997~98년도에, 『역사과학』에서는 이보다 앞선 1990~93년도에 ‘발해’ 논문의 비율이 ‘고구려’ 논문의 비율보다 월등히 높았다(그림 3 참고). 발해사에

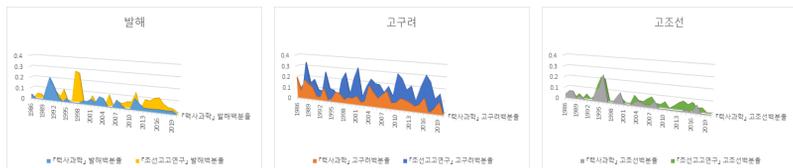
36) <표 4>에 나와 있듯이, 『역사과학』과 『조선고고연구』의 고구려 논문 비율은 각각 53%와 63%이다.

〈표 4〉 『조선고고연구』와 『역사과학』에 수록된, 제목에 ‘발해’, ‘고구려’, ‘고조선’이 포함된 ‘논문’의 수록 편수(1986년 이후)

제목 키워드	『조선고고연구』 ‘논문’ 수록 편수	『역사과학』 ‘논문’ 수록 편수
발해	58(20%)	49(26%)
고구려	179(63%)	102(53%)
고조선	49(17%)	40(21%)
합계	286	191



[그림 3] 『조선고고연구』(왼쪽)과 『역사과학』(오른쪽)에 수록된 발해사, 고구려사, 고조선사 논문의 비율과 그 통시적 변화 양상.



[그림 4] 『조선고고연구』와 『역사과학』에 수록된 발해사(왼쪽), 고구려사(가운데), 고조선사(오른쪽) 논문의 수록 비율과 그 통시적 변화 양상.

관심이 집중되었던 시기가 『조선고고연구』와 『역사과학』에서 달리 나타나는 이러한 양상은 [그림 4]를 보면 더 잘 확인된다. 즉, [그림 4]의 가운데에 있는 ‘고구려’ 그래프를 보면, 『조선고고연구』와 『역사과학』의 추이가 비슷하게 나타나는 반면, 왼쪽의 ‘발해’ 그래프에서는 비율이 가장 높게 나타나는 지점이 서로 다르게 확인되었다.

3.1.2.3. 『조선고고연구』와 『역사과학』에 수록된 발해사 논문의 주제 현황

『조선고고연구』와 『역사과학』에 수록된 발해사 논문들의 주제를 보면 각 학술지의 특성이 충실히 반영되어 있음을 알 수 있다. 즉, 『역사과학』에는 정치, 경제 등 발해사 전반에 대한 이해를 도모하기 위한 연구가 주로 진행된 반면, 『조선고고연구』에는 발해 주민의 삶의 죽음의 공간과 연관된 물질자료(성곽, 무덤)가 중요한 연구의 주제였다³⁷⁾.

한편, 발해사 연구에서 고구려, 고려 등과의 ‘계승’ 관계가 중요한 연구의 주제인 것으로 남한 학계에서는 인식되어 왔다³⁸⁾. 그런데 [그림 5]를 통해 알 수 있듯이, 『역사과학』에 수록된 ‘계승’을 다룬 논문의 비

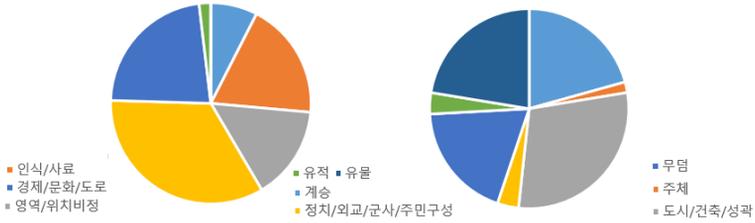
<표 5> 『조선고고연구』와 『역사과학』에 수록된 발해사 논문들의 공통 및 非 공통 주제

학술지	공통 주제				非공통 주제					
	계승	정치/외교/군사/주민구성	유적	유물	주체	도시/건축/성곽	무덤	인식/사료	영역/위치비정	경제/문화/도로
『조선고고연구』	12	2	2	13	1	17	11			
『역사과학』	4	18	1	0				10	8	12

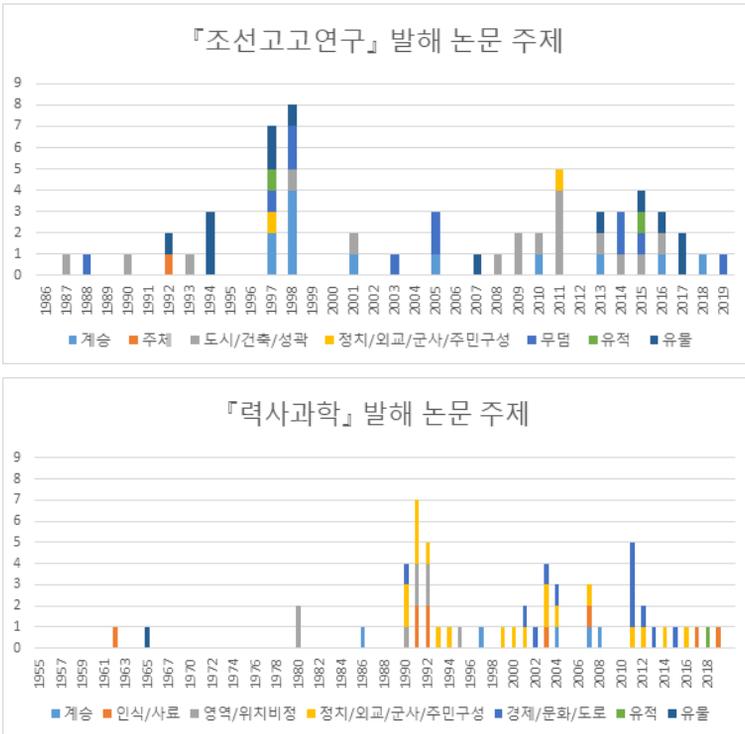
37) <표 4> 참고.

38) 예를 들어, 송기호(1990), p. 276.

『역사과학』 발해 논문 주제 『조선고고연구』 발해 논문 주제



[그림 5] 『조선고고연구』(왼쪽)와 『역사과학』(오른쪽)에 수록된 발해사 논문 주제 비율.



[그림 6] 『조선고고연구』(위)와 『역사과학』(아래)에 수록된 발해사 논문 주제의 통시적 분포 양상.

중이 높은 편은 아니다. 『조선고고연구』의 경우에는 ‘계승’에 관한 논문의 비중이 좀 더 높게 나타나기는 하나, [그림 6]에 잘 나와 있듯이 『조선고고연구』에 수록된 12편의 ‘계승’ 관련 논문 중 6편이 1997~98년도에 집중되어 있다. 2000년도 이후부터는 ‘계승’ 관련 논문이 주기적으로 보이기는 하지만, 이 주제가 역사학계이든 고고학계이든 북한의 발해사 담론을 ‘주도’하고 있다고 할 수는 없다.

3.2. 북한의 학술지 수록 논문에 대한 정성적/관계적 내용분석

3.2.1. 분석의 과정

『북한의 한국사 연구동향(1)』에 실린 이인철의 논고 「북한의 후기신라·발해사 연구동향」에서는 발해사를 다룬 논문 124편을 대상으로 계량적 분석을 진행하여 북한의 발해사 연구경향을 제시하였다.³⁹⁾ 이 연구에서는 정기적으로 간행되는 학술지뿐만 아니라, 『고고민속논문집』, 『역사과학논문집』, 『발해사연구논문집』 등에 수록된 논문도 함께 분석하여 총 14개의 주제를 확인하고 빈도분석을 통해 그 중요도를 파악하였다.⁴⁰⁾ 앞의 장에서 필자 역시, 비록 『조선고고연구』와 『역사과학』의 주제를 비교하고, 연구주제의 빈도를 통시적으로 접근하는 등 분석을 좀 더 세밀하게 진행했지만, 본질적으로는 이와 유사한 방식으로 북한의 발해사 연구주제에 대한 파악을 시도하였다.

그런데 주지하다시피 이렇게 주제 범주를 설정하는 행위에는 남한 연구자의 관점이 반영될 수밖에 없으며, 범주설정은 분석결과에 직접적인 영향을 미치게 된다. 따라서 연구주제의 동향을 파악한 이러한 기존의 성과들을 보완해줄 수 있는 새로운 접근이 필요한 실정인데, 자연

39) 이인철(2003), pp. 155-6.

40) 유적(발굴·성곽·건축·무덤)(31.5%), 기술·공예(15.3%), 계승성·주민구성(12%), 지리(9.7%), 대외관계(5.6%), 정치(5.6%), 역사(4.8%), 개설(4.8%), 경제(3.2%).

어 처리(Natural Language Processing) 기법을 활용한 넷마이너(NetMiner) 4.0 소프트웨어의 비정형 텍스트 데이터로부터 단어(노드)를 추출하는 기능을 이용하고, TF-IDF 지표를 이용해서 단어의 중요도를 파악하면 또 다른 방식으로 텍스트의 주제들을 추려낼 수가 있다.

다만 비정형 텍스트 데이터로부터 단어들이 자동적으로 추출된다고 하더라도, 연구자의 개입이 전혀 없는 것은 아니다. 우선 텍스트 데이터를 불러오는 과정에서 여러 가지 데이터 클리닝 행위들이 일어나는데, 1) 유의어 설정, 2) 지정어 설정, 3) 제외어 설정이 그것이다. ‘유의어 설정’은 유사한 의미를 가진 단어들이 하나의 단어로 추출되게 함으로써, 노드의 난립을 사전에 방지한다. 이러한 유의어 설정을 위해서는 연구자가 분석대상이 되는 텍스트를 검토한 다음에 유의어 사전을 미리 작성해야 한다. 본 분석에서 사용된 유의어 사전의 형태는 <표 6>에 제시되어 있다. ‘지정어 설정’은 고유명사, 복합명사 등의 추출을 위해 필요한 기능으로, 예를 들어 ‘정해공주묘’가 ‘정해’, ‘공주’, ‘묘’ 이 세 개의 개별 노드로 추출되는 것을 방지한다. 마지막으로, ‘제외어 설정’을 통해 지정된 단어들은 최종 결과 데이터에서 배제되는데, 본 분석의 경우에는 논문 제목에 흔히 포함되어 있지만 논문의 주제에 대한 정보를 담고 있지 않는 단어들을 제외어로 설정하였다. 그 예로, ‘고찰’, ‘연구’, ‘대하여’, ‘성격’, ‘형식’, ‘세기’ 등이 있다. 또한 앞서 언급했듯이 『력사과학』의 경우에는 ‘수령사’에 해당되는 논문들을 분석대상에서 배제했기 때문에 해당사항이 없었지만, 『조선고고연구』에는 종종 발해에 관한 논문 중에 “위대한 령도”와 같은 표현이 등장하는 제목들이 있어서 ‘위대한’, ‘령도’, ‘지도자’ 등도 제외어로 설정하였다.

위와 같은 과정을 통해 데이터 클리닝이 진행된 텍스트를 분석함에 있어서 두 가지 전략을 구사하였다. 우선, p. 208의 <표 3>에 제시된 북한의 대표 학술지에 수록된 모든 발해 논문의 제목을 분석의 대상으로 해서 시맨틱 네트워크를 구현하고 시각화를 진행하였다. 또한 학

〈표 6〉 시맨틱 네트워크 분석을 위해 진행된 데이터 클리닝 과정에서 사용된 유의어 사전(개념어 그룹 설정)

대표명칭	동일 그룹에 해당되는 개념어
관계	호상관계, 대일 관계, 국교 관계, 대일관계, 국교관계, 발해-당관계, 대외
계승	계승관계, 계승한 국가, 계승한 나라, 계승의식, 계승발전, 계승성
기와	마루밑막음기와, 마무리기와, 기와막새, 수기와막새, 마루수막새기와,련꽃무늬수기와막새
역참	역참제도, 역참포치, 간선도로역참포치
성곽	성, 수도성, 장성, 옹
발굴	발굴보고, 발굴중간보고, 조사
제철	제강, 철제품, 쇠
도자기	삼채, 자기, 질그릇
교통로	대일교통로, 력참로
말갈	말갈인, 흑수말갈
신라	후기신라
사관	력사관
발해	발해국
활	활촉
무덤	묘, 무덤떼
연해주	연해
고조선	조선

술지에 따라 주제의 차이가 있는지를 확인하기 위해 『조선고고연구』, 『력사과학』, 『민족(문화)유산』에 수록된 발해 논문에 대해 각각의 시맨틱 네트워크를 구현하고 시각화를 진행하였다. 마지막으로, 필요에 따라 특정 학술지의 고구려 및 고조선 논문에 대한 시맨틱 네트워크 구현 및 시각화를 통해 비교분석을 진행하였다.

3.2.2. 분석 결과

3.2.2.1. 북한 학술지 수록 발해 논문 주제들의 내용 클라우드

내용 클라우드에서는 텍스트 내 단어들의 상대적 중요성이 단어의 크기로 표현된다. 그런데 문제는 한 단어의 중요성이 월등히 높을 경우에는, 그 다음으로 중요도가 높은 단어들 간의 크기의 차이는 눈의 잘 띄지가 않는다. 본 분석의 경우에서는 각 논문의 제목이 하나의 ‘문서’로 인식되었는데, 연구의 틀을 짤 때 ‘발해’가 포함된 제목들만을 분석대상으로 삼은 만큼, ‘발해’라는 단어의 빈도와 그것이 포함된 문서의 빈도 모두 최고치일 수밖에 없다. 따라서 앞서 제시한 방식으로만 데이터 클리닝을 진행된 발해 논문을 대상으로 내용 클라우드를 생성했더니 [그림 7] 위쪽 이미지와 같은 결과가 나왔다. 이 경우에는 ‘발해’를 제외하고는 ‘고구려’, ‘무덤’ 정도만 부각되어 있고, 그 외의 단어들 사이의 상대적 중요성을 확인하기란 쉽지 않았다. 이런 경우에는 ‘로그 변환’ 등을 통해 척도의 격차를 좁히는 변환을 수행하곤 하는데, 본 연구에서는 데이터 클리닝 과정에서 ‘발해’도 제외어로 설정하여 다시 내용 클라우드를 생성하는 방법을 택하였다(그림 7, 아래).⁴¹⁾ 그 결과 ‘고구려’와 ‘무덤’ 다음으로 ‘계승’, ‘유적’, ‘관계’,⁴²⁾ ‘성곽’, ‘발굴’이 중요도가 높은 단어로 파악되었다.

이렇듯 내용 클라우드 생성 과정에서 ‘발해’를 제외어로 설정하는 것이 유의미한 결과 도출에 도움이 된다는 사실이 파악됨에 따라, 『조선고고연구』, 『력사과학』, 『민족문화유산』에 수록된 발해 논문에 대한 각각의 내용 클라우드를 생성하는 과정에서도 ‘발해’를 제외어로 설정하였다. 그 결과, [그림 7]에 제시된 세 개의 내용 클라우드가 생성되

41) 앞으로의 연구에서는 ‘로그 변환’ 작업을 수행하여 내용 클라우드를 생성하도록 하겠다.

42) 본 분석에서 ‘관계’는 ‘호상관계’, ‘대일관계’, ‘국교관계’, ‘발해-당관계’, ‘대일’, ‘대당’, ‘대신라’의 유의어로 설정되었다. <표 6> 유의어 사전의 2번 항목 참고.



[그림 7] 북한의 대표 학술지에 수록된 발해 논문 제목을 대상으로 생성한 ‘내용 클라우드’: ‘발해’ 키워드가 포함된 경우(위)와 ‘발해’를 제외어로 설정한 경우(아래).

있는데, 분석대상 문서의 개수가 가장 많았던(n=90) 『조선고고연구』의 경우에는 ‘무덤’의 중요도가 압도적으로 높고, 그 뒤를 이어 ‘고구려’와 ‘발굴’, 그리고 그 다음으로는 ‘유적’, ‘도자기’, ‘성곽’, ‘기와’가 높다는 것을 확인할 수 있다(그림 8, 왼쪽 위). ‘고구려’를 제외하고는, 이러한 양상은 고고학 중심의 학술지에서 충분히 예상할 수 있는 패턴이다. 한편, 『조선고고연구』에서는 나머지 두 학술지에 비해 ‘계승’이



[그림 8] 『조선고고연구』(왼쪽 위), 『민족문화유산』(오른쪽 위), 『역사과학』(왼쪽 아래) 수록 발해 논문 제목을 대상으로 생성한 ‘내용 클라우드’ 및 『역사과학』의 수록 고구려 논문 제목을 대상으로 생성한 ‘내용 클라우드’(오른쪽 아래).

부각되지는 않았지만, ‘고구려’의 중요도가 높게 나타났다는 사실은 『조선고고연구』의 발해 논문에서도 고구려와의 계승관계가 하나의 중요한 연구주제였음을 보여준다. 분석대상 문서의 개수가 가장 적었던(n=25) 『민족문화유산』의 경우에는 ‘고구려’, 그리고 그 다음으로 ‘계승’이 중요도가 가장 높은 단어였고, 나머지 두 학술지와는 차별되게 ‘풍습’의 중요도도 상대적으로 높았다(그림 8, 오른쪽 위).

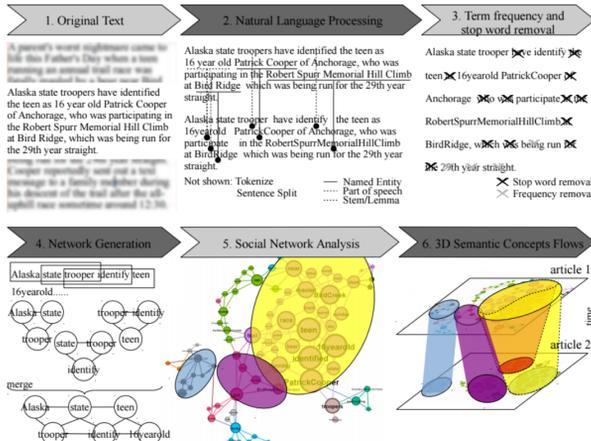
분석대상 문서의 개수가 두 번째로 많았던(n=56) 『역사과학』의 경우에는 ‘관계’가 중요도가 가장 높은 단어였고, 그 다음으로 ‘고구려’, ‘신라’, ‘계승’, ‘사관’, 그리고 그 뒤를 이어 ‘말갈’, ‘역참’, ‘국가’, ‘영역’이 상대적으로 중요도가 높은 것으로 파악되었다(그림 8, 왼쪽 위). 또한 『조선고고연구』에 비해 『역사과학』의 발해 논문들이 특정 주제에 편중되지 않고, 발해사를 다방면에서 접근하고 있음이 확인된다. 『조

선고고연구』의 경우에는 ‘발해사’보다는 ‘고고학’의 영향을 많이 받은 단어들이 내용 클라우드에서 부각된 것과 달리, 『역사과학』에서는 ‘발해사’ 관련 연구주제를 나타내는 단어들이 부각되었던 만큼, 『역사과학』에 한정하여 ‘발해’ 논문과 ‘고구려’ 논문의 내용 클라우드를 비교하는 것이 의미가 있다고 판단하고 후자도 생성하였다(그림 8, 오른쪽 아래). 그 결과, 『역사과학』의 고구려 논문(n=102)의 경우에도 ‘관계’와 ‘계승’의 중요도가 높다는 것을 확인할 수 있었다. 다만, 발해 논문들에서 ‘고구려’가 상당히 중요한 단어였던 것에 비해, 고구려의 논문들에서는 ‘발해’의 중요도가 상대적으로 낮았다. 한편, 흥미롭게도 『역사과학』의 고구려 논문들이 고고학의 성과들(‘무덤’, ‘성곽’)도 적극적으로 활용했음을 알 수 있다.

3.2.2.2. 북한 학술지 수록 발해 논문 주제들의 시맨틱 네트워크

북한의 발해 논문들을 대상으로 진행한 시맨틱 네트워크 분석에서, 앞서 다룬 결과들은 [그림 9]의 모식도에 제시된 3단계까지의 과정과 관련이 있었다. 이제부터는 [그림 9]의 4~5단계의 과정과 관련이 있는 결과들을 다루도록 하겠다.

시맨틱 네트워크를 생성하기 위해서는 두 가지 중요한 사항을 정해야 한다. 첫 번째로, 단어 간 네트워크를 생성할 때의 범위를 정해야 한다. 이것을 ‘윈도우 크기’(window size)라고 부르는데, 윈도우 크기를 최솟값인 ‘2’로 설정할 경우에는 나란히 등장하는 단어들끼리의 네트워크가 생성된다. [그림 9]의 ‘4. Network Generation’ 이미지는 윈도우 크기가 ‘3’으로 설정된 단어 간 네트워크를 나타내고 있다. 일반적으로 윈도우 크기가 아주 작을 경우에는 통사구조(syntactic structure)가 강조되고, 큰 경우에는 의미적 관련성(semantic relatedness)이 강조된다고 한다.⁴³⁾ 본 분석에서는 발해 논문 제목을 구성하는 단어들 간의 의미적 관련성 파악에 초점을 맞추고자 하니, 각 제목 당 모든 단어들 간의



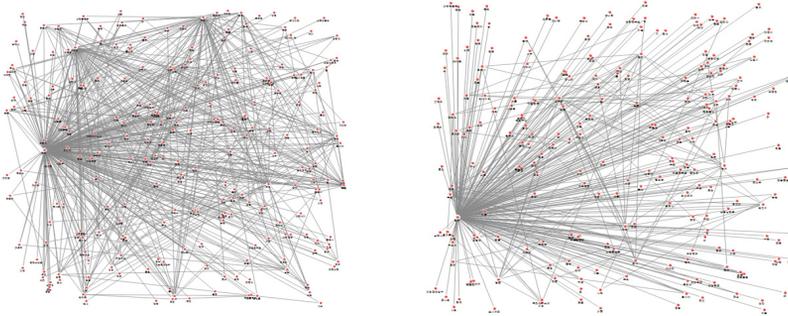
[그림 9] 시맨틱 네트워크 분석의 과정을 나타낸 모식도.⁴⁴⁾

네트워크가 생성될 수 있도록 윈도우 크기 값을 ‘10’으로 하였다. 두 번째로 정해야 하는 사항은 유의미한 단어 간 링크의 출현 빈도를 나타내는 ‘링크 빈도 임계치’(link frequency threshold)의 값이다. 이 값을 ‘1’로 설정하면 1번만 등장하는 링크의 단어들도 추출되는 만큼, 본 분석에서는 복수로 등장하는 링크들을 파악하기 위해 링크 빈도 임계치 값을 ‘2’ 이상으로 설정했다.

이러한 조건을 정하고 나서 넷마이너 4.0 소프트웨어를 이용해서 모든 발해 논문을 대상으로 단어들 간의 네트워크(word network)를 우선 생성하였다. 이때 링크 빈도 임계치 값을 ‘3’으로 설정하였다(그림 10, 오른쪽). 이는 그 값을 ‘2’로 설정했을 경우에는(그림 10, 왼쪽) 네트워

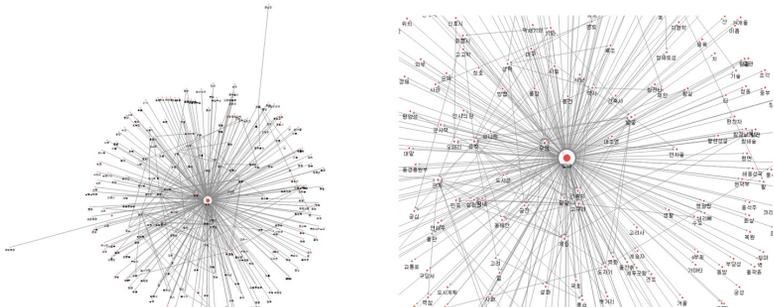
43) Levy, J. P., J. A. Bullinaria, & M. Patel (1999), “Explorations in the derivation of word co-occurrence statistics”, *South Pacific Journal of Psychology* 10(1), p. 104.

44) Cenek, M., R. Bulkow, E. Pak, L. Oyster, B. Ching, & A. Mulagada (2019), “Semantic Network Analysis Pipeline — Interactive Text Mining Framework for Exploration of Semantic Flows in Large Corpus of Text”, *Applied Sciences* 9, 5302, Figure 1 인용.



[그림 10] 발해 논문 제목을 대상으로, 링크 빈도 임계치 값을 각각 ‘2’(왼쪽)와 ‘3’(오른쪽)으로 설정했을 때 생성된 네트워크.

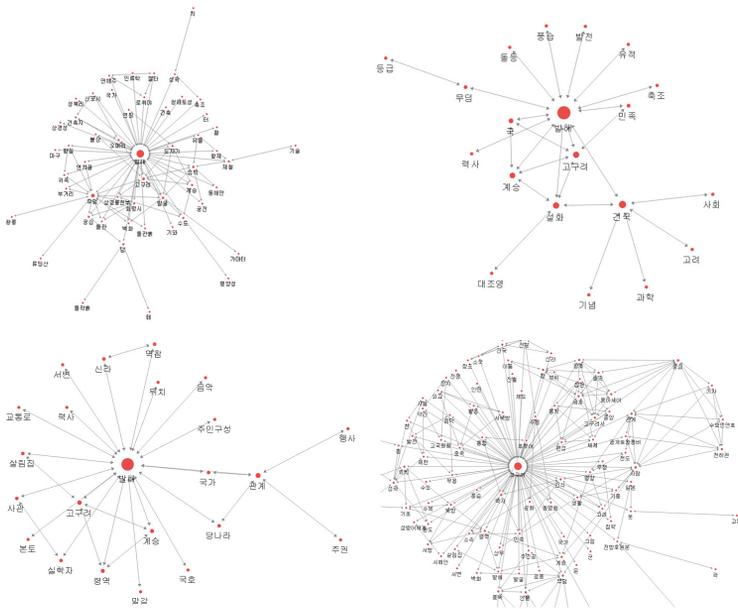
크가 너무나 많은 링크들로 구성되어 중요한 패턴들이 잘 드러나지 않았기 때문이다. 공기관계(共起關係, co-occurrence)를 표현한 이러한 네트워크는 발해 논문 제목에 등장하는 단어들 간의 의미적 관련성을 나타내는 데에는 유용하다. 하지만 북한 발해사 연구의 핵심 키워드(주제) 간의 관련성이 더 잘 드러나도록 할 필요가 있다고 판단되어, (링크 빈도 임계치 값을 ‘3’으로 설정하여 생성한) 네트워크를 구성하는 노드들의 연결중심성(degree centrality)도 분석하였다(그림 11). 그 결과 연



[그림 11] 발해 논문 제목 단어 네트워크에 대한 연결중심성 분석결과(왼쪽)와 연결성이 높은 단어들이 드러나도록 확대한 이미지(오른쪽).

결중심성이 높은 — 즉, 다수의 링크를 가진 — 노드들이 확인되었다. ‘고구려’, ‘무덤’, ‘계승’, ‘유적’, ‘관계’, ‘성곽’, ‘발굴’ 등이 해당되는데, 이 단어들은 앞서 생성된 내용 클라우드에서도 부각되었다.

한편으로 『조선고고연구』, 『력사과학』, 『민족문화유산』 별로로 발제 논문 제목에 대한 네트워크를 생성하였다. 이 경우에는 분석대상의 개수가 많지 않은 경우도 있고, 다양한 단어 간 연결성의 파악이 중요하다고 판단되어, 네트워크 생성을 위한 단어의 링크 빈도 임계치 값을 ‘2’로 설정하였다. 이렇게 생성된 3개의 네트워크를 대상으로 연결중심성 분석을 진행해서 얻은 결과가 [그림 12]에 제시되어 있다. 각 학



[그림 12] 『조선고고연구』(왼쪽 위), 『민족문화유산』(오른쪽 위), 『력사과학』(왼쪽 아래)의 발제 논문 제목의 네트워크에 대한 연결중심성 분석 결과 및 『력사과학』의 고구려 논문 제목의 네트워크에 대한 연결중심성 분석 결과 (오른쪽 아래).

술지 별로 발해 논문의 핵심 주제어들과 그 관계가 잘 드러난다는 것을 알 수 있다. 또한 『력사과학』의 고구려 논문 제목에 대해서도 동일한 방식으로 네트워크를 생성하고 연결중심성 분석을 진행하여(그림 12, 오른쪽 아래), 『력사과학』의 발해 및 고구려 핵심 주제어의 양상과 관계구조를 파악했다. 그 결과, 고구려 논문의 주제들이 더 다양할 뿐만 아니라, 그 관계구조가 더 복잡하다는 사실을 확인할 수 있었다.

4. 고찰

본 연구에서는 북한 학계의 발해사 연구동향에 대한 정보를 제공하는 자료 중 ‘다량의 균질적 데이터 세트’의 성격을 가진 ‘학술지 논문 제목’을 분석대상으로 삼아, 계량적 분석과 시맨틱 네트워크 분석을 진행하였다. 계량적 분석의 경우, ‘빈도분석’을 진행했다는 점에서는 기존의 작업과 유사한 측면이 있는데, 다만 데이터의 통시적 분포양상 표현과 맥락화에 주목했다는 점에서 차별화를 두었다. 시맨틱 네트워크 분석의 경우, 디지털 기술의 도움을 받아 생성된 빅데이터의 다양한 시각화와 분석을 통해 발해사 연구동향에 대한 새로운 시사점을 확보할 수 있었다. 본 장에서는 각각의 분석방법이 어떻게 발해사 연구경향에 관한 기존의 서술과 상호보완적인 관계를 맺을 수 있는지를 살펴보도록 하겠다.

4.1. 북한의 학술지 수록 논문에 대한 계량적/개념적 내용분석의 성과

4.1.1. 데이터의 통시적 표현으로 드러나는 것들

본 연구에서는 계량적 데이터 분석의 결과를 통시적으로 표현했다. 그 결과를 이병건의 논문에서⁴⁵⁾ 제시되었던 발해건축연구사의 분기설

정과 비교하면 이러한 접근의 성과가 잘 드러난다. 우선, 1980년대를 발굴자료가 현격하게 늘어난 ‘조사확장기’라고 규정하였는데, 여섯 개 학술지의 발해 논문 수록 편수의 통시적 변화 양상을 살펴본 이번 연구의 분석 결과(그림 2, 아래)는 이러한 발굴자료의 증가가 1986년 『조선고고연구』 학술지의 발행 시작과 관련 있음을 보여준다. 이병건의 논문에서는 ‘조사확장기’ 관련 인용 자료로 『조선고고연구』 논문들이 주를 이룬다는 점이 이러한 상관관계를 시사하고 있다. 한규철도 발해사 연구가 몇 년간의 공백이 있는 이후 1986년부터 다시 본격화되었다가, 1990년부터 발해사가 중요한 위치를 차지하게 되었다고 주장했는데,⁴⁶⁾ 본 연구를 통해 이러한 발해사 연구의 본격화·부흥과 『조선고고연구』 개시 사이의 상관관계가 명확하게 시각화된 셈이다.

이병건은 또한 2000~10년대를 ‘조사침체기·연구융합기’로 규정하고, “2000년대가 되면서 발해유적 발굴이 미미해지고 전체적으로 건축을 포함한 발해문화에 대한 연구가 미진”한 시기⁴⁷⁾라고 보았다. 그런데 이번 분석의 결과(그림 2; 위, 아래)에 의하면 1997~98년도 기간에 비하면 2000년대 초반의 발해 연구를 미진해 보일지 모르겠으나, 그 이후도 꾸준히 증가하였으며, 더 중요하게는 발해 논문 수록 학술지의 다변화가 나타났다.

한편, 계량적 데이터 분석의 결과를 통시적으로 표현했더니 발해사 연구에서 ‘계승’ 주제의 중요성에 대한 제대로 된 평가가 가능해졌다. 이인철은 2002년까지 출간된 124편의 발해 연구 논문을 주제별 검토한 결과, ‘계승’ 관련 논문 10편을 확인하였고⁴⁸⁾, 이번 연구에서 생성된 내용 클라우드와 단어 네트워크의 연결중심성 분석 결과에서도 ‘계승’

45) 이병건(2018), 앞의 글.

46) 한규철(1994), 앞의 글, p. 126.

47) 이병건(2018), 앞의 글, p. 110.

48) 이인철(2003), 앞의 글, p. 156.

이 주요 주제어도 등장한다. 그런데 발해 연구 주제의 분포 양상을 통시적으로 표현한 [그림 6]을 보면 『조선고고연구』의 경우, 12편의 ‘계승’ 주제 논문들 중 절반이 1997~98년 기간에 출간되었고, 그 이후로는 18년에 걸쳐 6편이, 그것도 한 해에 1편을 상회하지 않게 출간되었음을 알 수 있다. 이렇듯 데이터의 통시적 표현은 특히 연구동향 파악을 위해서는 반드시 요구되는 사항이어서, 본 연구에서는 계량적 분석 결과물의 표현에서 이 부분에 신경을 가장 많이 썼다. 그런데 정작 새로운 시도에 해당되었던 내용 클라우드 및 단어 네트워크의 경우에는 이러한 통시성이 표현되지 않았는데, 이 문제에 대해서는 아래에서 다시 언급하도록 하겠다.

4.1.2. 고고학계와 역사학계의 차이 확인

역사학계와 고고학계에서 진행된 발해사 연구가 서로 다른 주기와 서로 다른 궤적을 가지고 있음은 기존의 연구사 서술에서도 인지되었다. 다만, 이러한 ‘서로 다름’에 대한 인지는 장·절의 구성 방식 등을 통해 시사되었을 뿐이고,⁴⁹⁾ 계량적 분석의 결과물 제시에서는 이러한 차이가 반영되지 않았다. 본 연구에서는 이러한 기존 접근의 보완을 위해 『조선고고연구』와 『역사과학』을 구분하여 분석 결과를 표현하는 것을 기본 원칙으로 삼았다.

또한 두 학술지로 대변되는 두 학문분과의 발해사 연구동향의 차이를 파악하기 위해, 분석의 범주를 ‘발해’, ‘고구려’, ‘고조선’으로 세분화하기도 했다(그림 3, 그림 4). 그 결과, 발해사 연구가 고구려사 연구를 압도했던 시기가 『조선고고연구』에서는 1997~98년도이고, 『역사과학』에서는 1990~93년도임이 확인되었다. 각 학술지별로 해당 시기를 연구 주제의 통시적 분포 양상(그림 6)과 비교해서 살펴본 결과, 『역사과학』

49) 예를 들어, 이인철(2003), 앞의 글의 목차: “III. 북한학계의 발해사에 대한 문헌사학적 연구”와 “IV. 북한학계의 발해사에 대한 고고학적 연구 구분”.

의 피크(peak) 구간부터 정치/외교/군사/주민구성 관련 주제에 대한 연구가 본격화되고, 이러한 추세가 이후 지속적으로 이어졌음을 알 수 있다. 반면, 『조선고고연구』의 피크(peak) 구간은 앞서 언급한 ‘계승’ 주제 논문의 편중과 관련이 있다.

4.1.3. 북한 지성사의 흐름 속에서 연구동향의 맥락화

북한의 발해사 연구동향, 특히 논문의 빈도 및 주제에 대한 분석 결과는 북한 지성사의 큰 흐름 속에서 해석할 때 더 유의미한 시사점을 확보할 수 있다. 따라서 본 연구에서는 ‘발해’뿐만 아니라, ‘고구려’, ‘고조선’ 키워드 논문의 현황도 통시적으로 살펴보고, 세 범주의 데이터를 비교함으로써 몇 가지 흥미로운 사실들을 파악하였다.

본 연구에서 비교분석을 진행한 기간은 북한 사회과학원 고고학연구소에서 2009년에 제시한 북한 고고학의 학문적 흐름 7단계 중 다음의 단계들에 해당된다. “혁명발전의 요구에 맞게 우리 인민의 유구한 역사와 우수한 문화를 과학적으로 체계화하기 위한 사업”이 진행된 5기(1980년 10월~1989년 12월), “단군릉의 발굴과 5000년의 민족사를 새롭게 정립하기 위한 사업”이 진행된 6기(1990년 1월~1994년 12월), “선군시대의 요구에 맞게 우리 민족의 유구한 역사를 종합체계화하며 더욱 빛내기 위한 사업”이 진행된 7기(1995년 1월~현재).⁵⁰⁾

‘고구려’ 논문의 경우에는 단군릉의 발굴로 ‘고조선’ 논의가 활발히 진행된 6기 동안 일시적으로만 주춤했을 뿐, 지속적으로 일정 비율 이상 출간되었으며, 7기에 들어와서도 『조선고고연구』에는 매년 수록된 논문 편수의 평균 17%를 상회한다. 이와 달리, ‘고조선’ 논문은 제7기의 초반에 집중적으로 분포하고 있는데, 이는 6기 연구 성과물의 출간과 관련이 있는 듯하다.

50) 양시은(2016), 「최근 북한 고고학계의 고구려 연구 동향: 『조선고고연구』를 중심으로」, 『高句麗渤海研究』 56, 고구려발해학회, p. 87.

‘발해’ 논문의 경우에는 ‘고구려’ 및 ‘고조선’ 논문보다 중요한 비중을 차지했던 시기가 『조선고고연구』과 『력사과학』에서 서로 다르게 나타난다. 『력사과학』에서의 피크(peak)인 1990~93년도는 선행 연구에서도 주목을 받은 바 있는데, 한규철은 사회과학원 역사연구소 및 고고연구소의 소장들이 발해사에 관심을 가지고 지원을 한 점이나 인재의 영입과 같은 제도적 배려가 이러한 발해사 연구의 전성기를 가져온 것으로 보았다.⁵¹⁾ 이와 달리, 『조선고고연구』에서의 피크(peak)가 1997~98년도에 집중된 점은 우선 이 무렵에 『발해사연구논문집』 제2권, 『발해사문답집』, 『발해사연구』 시리즈 7권이 출간되고, 발해 관련 학술회의가 열리는 등⁵²⁾ 역사학계의 활발한 발해사 연구 활동과 관련이 있는 듯하다. 그러나 또 한편으로는 당시의 고구려와의 계승성을 강조해야 하는 분위기과도 연관이 있어 보인다. 앞서 언급했듯이, 『조선고고연구』에서 ‘계승’ 주제를 다룬 논문들 중 절반이 이 시기에 집중되어 있다. 또한 ‘계승’ 주제 논문의 숫자가 상대적으로 적은(n=4) 『력사과학』에도 이 시기에 발해가 단군조선과 고구려를 계승했다는 승성호의 논문이 게재되었다.⁵³⁾

4.2. 북한의 학술지 수록 논문에 대한 정성적/관계적 내용분석의 성과

4.2.1. 디지털 기술과 데이터 시각화의 가치

전통적인 방식의 연구사 서술을 보완해줄 수 있는 새로운 접근의 탐색이 본 연구의 출발 지점이었다. 이러한 관점은 “인문학의 본질로서 기능 해온 전통적 방식의 글쓰기는 그 역할 그대로 존중하되, 그것이 아

51) 한규철(1994), 앞의 글, pp. 126-127.

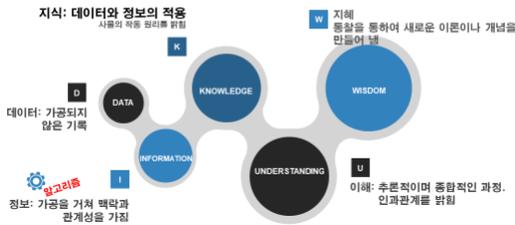
52) 이인철(2003), pp. 153-154 논저 목록 참고.

53) 승성호(1997), 「발해는 고구려의 영역을 계승한 나라」, 『력사과학』 1997-3, 사회과학출판사.

닌 다른 방식의 리터러시가 가능한지 가능하다면 그 구체적 형식이 무엇인지”를 타진해보고자 하는 디지털 인문학의 고민과⁵⁴⁾ 일맥상통한다.

필자의 경우에는 ‘전자 글쓰기’⁵⁵⁾에 능숙하지 않아 ‘자연어 처리 기술’의 도움을 받아 텍스트를 분석한 다음 그 결과에 대한 데이터 시각화를 진행하였다. 그 결과가 바로 앞서 제시한 ‘내용 클라우드’ 이미지들인데, 주지하다시피 통계정보를 그래프로 보여주는 전통적인 방식과는 달리, ‘데이터’의 요약이 이루어졌지만 한편으로는 기존보다 훨씬 더 많은 데이터를 한 눈에 살펴볼 수 있게 되었다.⁵⁶⁾ ‘내용 클라우드’로 시각화된 데이터는 그 자체로도 많은 ‘정보’를 제공하였고, 또 한편으로는 학술지 간, 주제 간 비교분석도 훨씬 더 직관적으로 이루어질 수 있도록 함으로써 복한 고고학계의 학문 생태계에 대한 ‘지식’의 생성에도 중요한 기여를 했다(‘데이터’, ‘정보’, ‘지식’의 차이는 [그림 13]에 제시되어 있다).

DIKUW hierarchy



[그림 13] ‘데이터-정보-지식-이해-지혜’의 프레임워크.⁵⁷⁾

54) 류인태(2020), 「디지털 인문학은 인문학이다」, 『인문논총』, 서울대 인문학연구원 77(3), p. 402.

55) 류인태(2020), p. 379에서 ‘전자 글쓰기’에 대해 설명하고 있다.

56) 최광선 · 함영경 · 김선호(2013), 『빅데이터 시각화』, 서울: 한국컴퓨터정보학회.

57) Ackoff, R. (1989), “From Data to Wisdom”, *Journal of Applied Systems Analysis*

그런데 ‘자연어 처리 기술’이라는 지름길을 통해 ‘데이터 시각화’의 세계로 입문하게 된 필자는 역설적으로 ‘전자 글쓰기’ 기술의 필요성을 절감하게 되었다. 왜냐하면 데이터 시각화의 필요성과 무한한 가능성에 대해 알게 되었지만, 한편으로는 ‘자연어 처리 기술’에만 의존해서는 그것의 가치를 제대로 활용할 수 없다는 사실을 인지하게 되었기 때문이다. 일례로 『『지암일기』 디지털 인문학 연구(URL: <http://jiam-diary.info/>)』에 활용된 다양한 데이터 시각화의 사례들은 연구자의 ‘전자 글쓰기’를 반드시 필요로 한다. 특히, 최근에 와서는 ‘플로우 데이터(flow data)’의 시각화를 위한 다양한 도구들이 제공되고 있는데, 통시적 접근에 유용하게 사용될 수 있는 이러한 도구를 사용하기 위한 선행조건이 연구자의 ‘전자 글쓰기’가 되겠다.

물론 그렇다고 ‘자연어 처리 기술’을 이용한 주제어의 자동 추출의 의미를 폄하해서도 안 될 것이다. 이러한 기술이 없었다면 필자는 단어의 추출이나 링크의 설정에 상당한 시간을 투자해야 했을 것이다. 또한 사람의 판단이 아닌 알고리즘에 의한 단어의 자동 추출은 무엇보다도 객관성을 담보하며, 데이터 세트가 공개되었다면 과정의 재현도 가능한 만큼 논쟁의 여지가 있는 본 연구의 주제에 — 즉, 남한 학자에 의한 북한 학계의 연구동향 연구 — 특히 적합하다. 이러한 측면에서 볼 때, ‘자연어 처리 기술’을 활용한 이번 분석이 통일 이후를 대비한 공통기반의 구축을 추구하는 ‘통일기반구축사업’의 취지와 잘 맞아 떨어진다고 판단된다.

4.2.2. 네트워크가 보여주는 것들

데이터 시각화를 지도학에 최초로 도입한 연구자 중 하나인 디비아즈(Dibiase)에 의하면, 지도는 ‘시각적 사고’의 수행에 중요한 역할을 하

16의 내용을 재구성한 인포그래픽으로, 허정원이 제공함.

며, 이것이 ‘시각적 커뮤니케이션’ 못지않게 중요한 지도의 기능이다.⁵⁸⁾ 이와 마찬가지로 본 연구에서 제시된 ‘단어 네트워크에 대한 연결중심성 분석 결과’의 데이터 시각화 사례는 단순히 분석 결과의 전달에 머무는 것이 아니라, 새로운 연구 가능성의 발견을 가능하게 하는 도구가 된다. 즉, 발해 논문의 키워드들이 노드와 링크의 형태로 재구성되면서 새로운 연구의 방향이 제시되었다.

중심연결성이 높은 노드들은 사실 기존의 분석 결과에 입각해서 충분히 예상할 수 있었다. 하지만 그보다 중심연결성이 한 단계 낮은 노드들이 다른 노드들과 링크된 양상은 그 자체로서 새로운 방식으로 연구동향을 제시한 것일 뿐만 아니라, 새로운 시사점도 제공하였다. 또한, 『역사과학』의 발해 및 고구려 논문 주제들이 보여준 네트워크의 차이는 새로운 연구 영역의 확장을 제시했다고 평가할 수 있다.

4.2.3. 통시성 확보를 위한 앞으로의 노력

본 연구에서 진행한 정성적/관계적 내용분석은 여러 가지 가능성을 보여주었지만, 또 한편으로는 분명한 한계를 가지고 있는데, 그것을 바로 통시성이 고려되지 않았다는 것이다. 앞서 살펴 본 ‘시맨틱 네트워크 분석의 과정을 나타낸 모식도’(그림 9)의 6단계가 ‘3D Semantic Concept Flows’이라는 점이 보여주듯이, 데이터 시각화를 활용해서 통시적 변화양상도 당연히 나타낼 수 있다.⁵⁹⁾ 그런데 이러한 작업을 진행하기 위해서는 관련 시각화 소프트웨어뿐만 아니라, 어느 수준 이상의 데이터 처리 기술이 요구된다. 빅데이터의 통시성을 시각화하는 기

58) 정인철(2015), 「한국 지도학 발달사」, 『지도학의 발달과 패러다임의 변화』, 서울: 진한엠앤비, p. 23.

59) 통시성을 잘 나타내는 데이터 시각화의 도구와 사례들은 김지우·이건학(2017), 「웹기반 데이터 시각화 도구를 활용한 플로우 데이터의 지리적 시각화 기법 탐색」, 『한국지도학회지』 17(1), 한국지도학회에 잘 소개 되어있다.

술이 이미 존재하고 널리 활용되고 있는 상황인 만큼, 통시성이 연구와 불가분의 관계를 맺고 있는 고고학 및 역사학 분야의 연구자들은 ‘전자 글쓰기’ 실력을 향상시켜 그러한 기술을 활용하는 방안을 모색해야 할 것이다.

5. 결론

전통적인 방식으로 연구동향을 서술한 논문은 세상에 공개되고 얼마 지나지 않아 ‘유통기한’을 다하게 된다. 본 연구도 지면으로만 유통되면 그러한 운명을 피하기 어려울 것이다. 그러나 연구 결과물이 지면에만 갇혀있지 않고 온라인상으로 존재하게 된다면 그 생명을 연장할 수 있을 것이다. 즉, 필자는 이번 분석에 사용한 모든 데이터를 공개하여, 원 데이터가 꾸준히 업데이트 되고, 또한 누구나 새로운 문제의식에 입각해서 새로운 연구의 틀을 짜서 북한 학계의 발해사 연구동향에 대한 새로운 정보와 지식과 지혜를 창출해 나가기를 기대한다. 이러한 포부가 원대하기는 하나, 불가능한 것은 아님을 보여주는 것이 앞서 언급한 『지암일기』 디지털 인문학 연구(URL: <http://jiamdiary.info/>) 프로젝트와 ‘Signs at 40’(<http://signsat40.signsjournal.org/>) 프로젝트의 사례이다.

이렇듯 온라인 플랫폼을 통한 연구대상 자료와 방법론의 공유를 바라는 이유는 그것이야 말로 본 연구에서 구사한 접근과 방법론의 확산을 위해 유용하다고 판단하기 때문이다. ‘전자 글쓰기’ 실력의 향상과 데이터 시각화의 적극적인 활용은 인문학의 외연 확장에 중요한 기여를 할 수 있으며, 이러한 연구에 입문하는 문턱을 낮추면 패러다임의 전환을 앞당길 수 있다고 생각한다.

물론 디지털 데이터 분석 방법론이 더 많은 연구자들에게 유용한 것

으로 인식되기 위해서는, 본 연구와 같은 연구동향에 대한 메타분석뿐만 아니라, 각 분야의 주된 관심의 대상이 되고 있는 자료에 대한 분석을 진행해야 할 것이다. 다행이 이러한 노력은 조금씩 진행되고 있는 추세라서,⁶⁰⁾ 앞으로 인문학적 데이터의 디지털 분석과 시각화는 더 활발히 이루어질 것을 기대할 수 있을 것이다.

60) 앞서 언급한 ‘『지암일기』 디지털 인문학 연구(URL: <http://jiamdiary.info/>)’ 프로젝트뿐만 아니라, 2020년 10월에 발표된 다음과 같은 사례도 있다: 허수(2020), 「텍스트 마이닝을 활용한 개벽의 논조(論調) 분석」, 『개벽의 백년 백년의 개벽 — 개벽으로 다시 여는 매체인문학 연구의 새 지평』 — 성균관대학교 대동문화연구원 한림대학교 한림과학원 아시아문화연구소 『개벽』 창간 백주년 기념 공동기획 학술대회 발표 자료집, 성균관대학교 대동문화연구원.

참고문헌

【논 저】

- 국사편찬위원회(2001), 『북한 역사학 논저 목록(하)』, 서울: 국사편찬위원회.
- 김지우·이건학(2017), 「웹기반 데이터 시각화 도구를 활용한 플로우 데이터의 지리적 시각화 기법 탐색」, 『한국지도학회지』 17(1), 한국지도학회.
- 류인태(2020), 「디지털 인문학은 인문학이다」, 『인문논총』 77(3), 서울대 인문학연구원.
- 리정기(1967), 「청해토성 및 교성리토성 답사보고」, 『고고민속』 1967-4, 사회과학원 고고학연구소.
- 리준걸(1984), 「새로 발굴된 24개돌」, 『력사과학』 1984-2, 과학백과사전출판사.
- 송기호(2012), 「발해의 문화와 발해사 인식의 변천」, 『발해(신편 한국사 10)』, 서울: 국사편찬위원회.
- _____(1991), 「北韓의 渤海史·統一新羅史 研究」, 『北韓의 古代史研究』, 서울: 一潮閣.
- _____(1990), 「북한의 발해 고고학과 《발해문화》」, 『역사와 현실』 3, 한국역사학회.
- 승성호(1997), 「발해는 고구려의 영역을 계승한 나라」, 『력사과학』 1997-3, 사회과학출판사.
- 양시은(2016), 「최근 북한 고고학계의 고구려 연구 동향: 『조선고고연구』를 중심으로」, 『高句麗渤海研究』 56, 고구려발해학회.
- 이동휘(2020), 「북한 경내의 발해유적 발굴조사성과와 그 의의 — 부거리와 북청 일대 —」. 『역사와 세계』 57(1), 효원사학회.
- 이병건(2018), 「북한의 발해건축 관련 연구와 향후 우리의 과제」, 『高句麗渤海研究』 61, 고구려발해학회.
- _____(2007), 「발해건축에 관한 북한의 연구성과」, 『高句麗渤海研究』 26, 고구려발해학회.
- 이인철(2003), 「북한의 후기신라·발해사 연구동향」, 『북한의 한국사 연구동양(1)』, 서울: 국사편찬위원회.

- 정경일·王寧(2019), 「2019년 북한학계의 고구려 및 발해사 연구 목록」, 『高句麗渤海研究』 65, 고구려발해학회.
- 정경일·김명(2019), 「2019년 북한학계의 고구려 및 발해사 연구 목록」, 『高句麗渤海研究』 65, 고구려발해학회.
- 정인철(2015), 「한국 지도학 발달사」, 『지도학의 발달과 패러다임의 변화』, 서울: 진한엠앤비.
- 최광선·함영경·김선호(2013), 『빅데이터 시각화』, 한국컴퓨터정보학회.
- 한규철(1994), 「북한의 발해사 연구」, 『北韓의 古代史研究와 성과』, 서울: 대륙연구소.
- 허수(2020), 「텍스트 마이닝을 활용한 개벽의 논조(論調) 분석」, 『개벽의 백년 백년의 개벽 — 개벽으로 다시 여는 매체인문학 연구의 새 지평』 — 성균관대학교 대동문화연구원 한림대학교 한림과학원 아시아문화연구소 『개벽』 창간 백주년 기념 공동기획 학술대회 발표 자료집, 서울: 성균관대학교 대동문화연구원.
- Ackoff, R. (1989), “From Data to Wisdom”, *Journal of Applied Systems Analysis* 16.
- Berelson, B. (1952), *Content Analysis in Communication Research*, Glencoe, Ill.: Free Press.
- Binsbergen, J. (2013), “Quantitative Content Analysis”, <https://www.digitalmethods.net/MoM/QuantContentAnalysis>
- Cenek, M., R. Bulkow, E. Pak, L. Oyster, B. Ching, & A. Mulagada (2019), “Semantic Network Analysis Pipeline — InteractiveText Mining Framework for Exploration of Semantic Flows in Large Corpus of Text”, *Applied Sciences* 9.
- Choi, S. H., H. J. Seo, & Y. S. Kim (2016), “Analysis of the Research Trends of the Korean Journal of Educational Research Using Network Text Analysis”, *International Journal of Software Engineering and Its Applications* 10(12).
- Cidell, J. (2010), “Content clouds as exploratory qualitative data analysis”, *Area* 42(4).
- Cyram (2017), 『NetMiner 4.0 Semantic Network Analysis 매뉴얼』.

- Goodman, R. A., S. B. Thacker, & P. Z. Siegel (2001), "What's in a Title? A Descriptive Study of Article Titles in Peer-Reviewed Medical Journals", *Science Editor* 24(3).
- Graham, L., W. Lepenies & P. Weingart(eds.) (1983), *Functions and Uses of Disciplinary Histories*, Dordrecht: D. Reidel Publishing Company.
- Holsti, O. R. (1969), "Content Analysis", *Content Analysis for the Social Sciences and Humanities*, Boston: Addison-Wesley.
- Hsieh, H.-F. & S. E. Shannon (2005), "Three approaches to qualitative content analysis", *Qualitative Health Research* 15(9).
- Hudson, J. (2016), "Analysis of the titles of papers submitted to the UK REF in 2014: authors, disciplines, and stylistic details", *Scientometrics* 109.
- Jackson, A., R. Harris, L. Hepple, A. Hoare, R. Johnston, K. Jones, & P. Plummer (2006), "Geography's changing lexicon: measuring disciplinary change in Anglophone human geography through journal content analysis", *Geoforum* 37.
- Jamali, H. K. & Nikzad, M. (2011), "Article title type and its relation with the number of downloads and citations", *Scientometrics* 88(2).
- Krippendorff, K. (2004), *Content Analysis: An Introduction to Its Methodology* (2nd ed.), Thousand Oaks, CA.: Sage.
- Levy, J. P., J. A. Bullinaria, & M. Patel (1999), "Explorations in the derivation of word co-occurrence statistics", *South Pacific Journal of Psychology* 10(1).
- Patton, M. Q. (2002), *Qualitative Research and Evaluation Methods*, Thousand Oaks, CA: Sage.
- Prasad, B. D. (2019), "Qualitative Content Analysis: Why is it Still a Path Less Taken?" *Forum: Qualitative Social Research* 20(3).

원고 접수일: 2020년 11월 1일

심사 완료일: 2020년 11월 10일

게재 확정일: 2020년 11월 11일

ABSTRACT

Content Analysis of North Korean Research Journal Data:
a New Way of Writing Histories on Balhae Research Undertaken
in North Korea

Ko, Ilhong*

Previous attempts to trace North Korean research trends have relied heavily on qualitative analysis, resulting in subjective narratives, and limited attempts at quantitative analysis have been one-dimensional in nature. In order to address this need for the diversification of methods for the study of North Korean Balhae research, ‘digital data analysis’ methods widely used in the social sciences were adopted. This paper presents the results of qualitative and quantitative content analysis undertaken on the titles of articles on Balhae published in six North Korean research journals. Descriptive statistical analysis was used, as was Semantic Network Analysis. The results of the former revealed diachronic insights on the publication of Balhae articles, and comparative analysis was undertaken on different journals. The latter involved the use of NetMiner 4.0 software to produce a semantic network, to create ‘content clouds’, and to undertake degree centrality analysis. The results of the this study contributed to a better under-

* HK Research Professor, Seoul National University Asia Center

taking of North Korean Balhae research trends, presented a way of utilizing data on North Korean research journals accumulated through DB construction projects, and provided an example of applying digital data analysis techniques to humanities research.

