

국내 번역연구의 어제와 오늘: 텍스트 마이닝 분석 — 『번역학연구』를 중심으로*

이 창 수
(한국외국어대)

1. 들어가는 말

국내에서 번역에 대한 학문적 연구가 본격적으로 시작된 것은 1990년대 말~2000년대 초에 3종의 전문학술지가 차례로 등장한 이후부터라고 할 수 있다. 1997년 한국외국어대학교 통번역연구소가 『통번역학연구』를 창간하였고 다음 해인 1998년에 한국국제회의통역학회(KSCI)가 『국제회의 통역과 번역』(현 학술지명 『통역과 번역』)을 출범시켰으며, 바로 뒤이어 2000년에 한국번역학회에서 『번역학연구』를 창간하였다. 이들 학술지가 창간된 후 이제 20년 가까운 시간이 흘렀기 때문에 그동안 번역 연구의 주제와 방법에서 큰 변화가 있었을 것으로 보인다. 이 같은 배경에서 본 연구는 2000년대 초의 학술지 발간 초기와 최근에 발표된 번역 연구 논문들 간에 연구 주제나 방법 면에서 어떤 차이

* 본 연구는 2018년도 한국외국어대학교 교내연구비 지원을 받아 작성되었음.

가 있는 지를 분석해보는 것을 목적으로 한다.

어떤 학문 분야의 연구 동향을 분석하는 데는 여러 방법이 있을 것이다. 일반적으로는 분석 논문을 수집하여 이를 내용이나 주제에 따라 코딩-분류한 후 그 결과를 통계 분석하는 방법을 생각할 수 있다. 이 같은 전통적인 콘텐츠 분석기법은 세밀한 내용을 분석할 수 있는 장점이 있지만 데이터 분류, 코딩, 해석 과정에서 연구자의 주관이 개입될 여지가 있다(베일리(Bailey) 1994: 318). 무엇보다 주제 분야를 미리 정해놓고 그 틀에 맞춰 분류하는 특징이 있다. 이와 비교하여 최근에는 콘텐츠 분석에 데이터 마이닝이나 텍스트 마이닝과 같은 컴퓨터를 활용한 분석기법이 많이 시도되고 있다(cf. 샬리(Shali) 2017). 가령, 유효현과 신세인(2015)은 의학교육 학술지인 『한국의학교육』에 15년간 실린 274편의 논문의 주제어를 대상으로 사회관계망분석 상용프로그램인 넷마이너(NetMiner)를 사용하여 주제군을 네트워크로 도식화하는 연구를 진행하였다. 본 연구에서는 이와 같은 상용프로그램을 사용하는 대신에 R언어를 사용한 텍스트 마이닝 기법과 게피(Gephi)라는 네트워크 생성 프로그램을 활용하여 번역학 분야의 연구 동향 변화를 분석해보고자 한다. 이 같은 분석방법은 논문의 내용을 상세하게 분류·정리하기보다는 연구의 전반적인 주제 구조를 파악하는데 효과적이다. 따라서 본 연구도 2000년대 초와 현재 사이에 국내 번역연구의 큰 흐름이 어떻게 변했는가를 알아보는데 그 목적을 두고 있다.

분석에 사용된 데이터는 『번역학연구』에 발표된 논문으로 한정하였다. 이는 동 학술지가 출범 초부터 연 4회 학술지를 발간하였기 때문에 학술지 출범 초기에 대하여 충분한 양의 번역관련 논문을 확보하는 것이 가능했기 때문이다. 이에 반하여 『통번역학연구』와 『번역과 통역』은 발간 초기에 발간 회수가 연 1, 2회로 제한적이었고 통역 분야 논문이 많아 번역연구에 국한해 시대별 비교를 하는데 충분한 분석 데이터 확보가 불가능했다.

본 연구를 통해 학문의 특성상 원문과 번역문 분석에 치중될 수밖에 없는 번역연구의 연구 지평을 확대하는 효과를 기대한다. 특히 빅데이터, 인공지능, 기계번역이 화두로 등장하고 있는 요즘 이러한 흐름과 맥을 같이하는 컴퓨터 활용 분석기법의 도입은 번역학을 한 단계 발전시키는데 중요한 역할을 할 것으로 기대된다.

2. 국내 번역연구 동향 연구

국내 학술지에서 번역학에 대한 연구 동향을 다룬 초기 논문은 김지원(2000)이 있지만 이는 해외의 연구 동향을 소개한 논문이다. 국내 번역 연구가 본격적으로 시작된 것이 1998~2000년 초라는 점을 고려하면 2000년 초에 국내 번역연구 동향을 분석한 논문이 없는 것은 당연해 보인다. 그 외에 번역연구 동향을 주제로 한 국내 논문은 김혜림(2012), 서유경(2013)이 있는데 이는 각각 중국과 러시아의 통·번역 연구 동향을 소개한 논문이다. 국내 번역연구 동향을 본격적으로 분석한 최초의 논문은 정호정(2013)과 김혜림(2013)으로 두 논문이 모두 2013년도에 발간되었다는 점이 특이하다. 이는 국내 번역연구의 동향을 분석하기 위한 충분한 연구 성과가 축적되는데 15년 정도의 기간이 걸렸다는 점을 시사한다.

정호정(2013)은 국제학술지인 *Meta* 및 *The Translator*와 국내 학술지인 『통역과 번역』 및 『번역학연구』를 대상으로 1999년에서 2012년까지 번역관련 논문을 취합하여 연구논문수와 함께 7개의 대주제별로 연구 내용을 분류 분석하였다. 국내 학술지 분석 결과를 보면 양 학술지 전체 논문의 1/4을 차지할 정도로 번역이론에 관한 논문이 가장 큰 비중을 차지하였다. 또한 비문학(전문번역 포함)과 문학번역으로 나눌 경우 2004년을 기점으로 『통역과 번역』에선 문학번역이 전문번역보다 높게 나타난 반면 『번역학연구』에선 반대 현상이 목격되었다.

번역분야에만 국한했던 정호정(2013)과 달리 김혜림(2013)은 2013년까지 발간된 국내 통번역관련 연구 논문 410편을 분석대상으로 삼아, 제목, 초록, 키워드에 기초하여 11개 주제별로 분류하였다. 그 결과 연구 주제면에서 텍스트분석(20.8%), 언어학적 관점(14.6%), 교육(14%) 등이 상위를 차지하였다.

두 논문 이후에는 특정 번역 연구 분야에 대한 동향 분석 논문들이 등장하였는데 영상번역의 연구 동향을 다룬 조성은(2014), 코퍼스를 활용한 연구 동향을 다룬 최희경(2015), 기계번역 연구 동향을 분석한 이견일과 이종혁(2015) 등이 있으며, 가장 최근에는 한현희(2017)가 통번역 교육 분야의 연구 동향을 분석하였다.

이상 연구들의 공통점은 연구 내용을 주제별로 분류하기 위하여 특정한 수의 범주를 설정하고 이에 따라 논문을 분류하였다는 점이다. 이와 달리 본 연구

에서는 특정한 범주를 사전에 설정하지 않고 분석 논문의 초록을 분석 코퍼스로 하여 텍스트 마이닝 기법을 활용하여 전체 코퍼스에서 드러나는 주제 구조와 변화의 큰 흐름을 분석하는 방식을 택하였다.

3. 분석 데이터 및 방법

본 연구에서는 앞서 언급하였듯이 국내 번역학 연구의 초기와 현재를 비교 분석하기 위하여 번역학 전문 학술지인 『번역학연구』를 대상으로 학술지 발간 초기인 2000년 창간호에서부터 2003년까지 발간된 46편의 논문과 최근 2015년에서 2017년 1호까지 발간된 75편의 논문 등 총 121편 논문의 영문 초록을 분석 데이터로 선택하였다. 논문의 편의상 초기 논문코퍼스를 JTS1이라고 하고 후기 코퍼스를 JTS2라고 부르기로 한다. 양 기간 중 통역을 주제로 한 논문은 분석에서 제외하였으나 통역과 번역을 모두 다룬 논문은 데이터에 포함시켰다. 후자의 경우 대부분은 통번역 교육과 관련된 논문이었다. 또한 JTS1에 해당하는 논문 중 학술지 초기에 영문 초록이 없는 소수의 논문도 분석에서 제외하였다.

이 같이 수집된 JTS1과 JTS2는 제목과 저자명 키워드 등을 제외하고 초록 본문만 각각 하나의 문서 파일로 통합하였다. 각 문서 파일에서 각 초록은 문단으로 구분되도록 하였다. 단어 수를 기준으로 JTS1의 크기는 12,003, JTS2는 14,461였다. JTS1에 비하여 JTS2에 포함된 논문 수가 29편 더 많기 때문에 JTS2의 어휘 수가 높은 것은 당연하지만 그 차이가 예상보다 크지 않은 것은 학술지 초기 논문에 해당하는 JTS1의 초록이 상대적으로 길기 때문이다.

이렇게 만들어진 JTS1과 JTS2 두 개의 문서파일을 R에 탑재한 후에 다음과 같은 일련의 전처리 작업을 진행하였다.

- 따옴표, 구두점, 숫자, 문서 앞과 뒤의 공백 등 분석에 불필요한 문자를 제거하였다.
- tm패키지의 stopwords 목록을 사용하여 전치사, 대명사 등 내용과 연관성이 없는 소위 기능어휘들을 제거하였다.
- 이 외에 stopwords 목록에는 포함되어 있지 않지만 내용분석에 불필요한 어

회 중 고빈도 어휘를 분석에서 제외하였다. 예를 들면, can, may같은 조동사, always, also, often과 같은 부사어, seem, make, think와 같은 일반 동사, one, two, three와 같은 숫자어휘, among, alike와 같은 전치사 등이 있다. 같은 맥락에서 글자 수가 3개 이하인 단어도 분석 데이터에서 삭제하였다. 이들 어휘도 거의 전부가 내용과 상관없는 것들이다.

- 일회성 주제의 영향을 배제하기 위하여 빈도수 5 이하인 어휘를 제거하였다.

이와 같은 전처리 과정을 거친 두 문서를 R의 tm 패키지를 활용하여 텍스트 마이닝 작업에 필요한 TDM(term-document matrix)으로 변환하였다. TDM이란 그림 1에서 볼 수 있듯이 각 코퍼스 내의 문서 이름(P1, P2, P3...)을 행, 개별 어휘를 열로 하여 각 문서별 어휘 빈도수를 표시한 행렬도로 텍스트 마이닝의 가장 기본이 되는 데이터 형태이다. 이와 같이 추출된 TDM을 기반으로 1차 어휘빈도 분석을 시행하였다. 특히 JTS1과 JTS2에서 상대적으로 비중이 큰 어휘를 시각적으로 쉽게 비교 분석하기 위하여 비교 워드클라우드 분석을 시행하였다.

〈그림 1〉 JTS1 TDM의 예

	A	B	C	D	E	F	G	H	I
1		translation	language	text	translator	word	study	difference	target
2	P1	20	0	4	2	0	4	0	1
3	P2	13	1	1	0	2	1	0	2
4	P3	5	0	0	1	0	1	0	0
5	P4	3	5	1	4	0	0	6	3
6	P5	6	1	2	0	3	0	2	0
7	P6	3	1	0	0	0	4	0	0
8	P7	0	2	0	2	0	0	2	0
9	P8	18	9	1	0	0	5	0	2
10	P9	3	5	0	0	2	1	0	0
11	P10	2	2	0	0	0	4	3	0
12	P11	0	1	5	1	2	0	2	1
13	P12	10	1	1	2	2	0	0	2
14	P13	7	6	2	0	0	2	1	4

다음에는 이와 같은 어휘가 서로 어떤 관계를 갖고 있는가를 분석하기 위하여 TDM을 각 초록 내에서 특정 어휘가 다른 어휘와 같이 발생하는 빈도를 표시한 상호연관성 행렬도(co-occurrence matrix)로 변환한 후 이를 다시 네트워크 데이터로 바꿔 네트워크 분석을 실시하였다. 이러한 네트워크 분석은 트위터나 페이스북과 같은 사회관계망을 통해 연결된 사람들의 관계를 분석하는 기

법으로 이를 어휘 간 관계 분석에 적용한 것이다(Garten 2010; Wiedemann 2016). 본 연구에서 네트워크 분석은 게피(Gephi, <https://gephi.org/>) 오픈 소스 프로그램을 사용하였다.

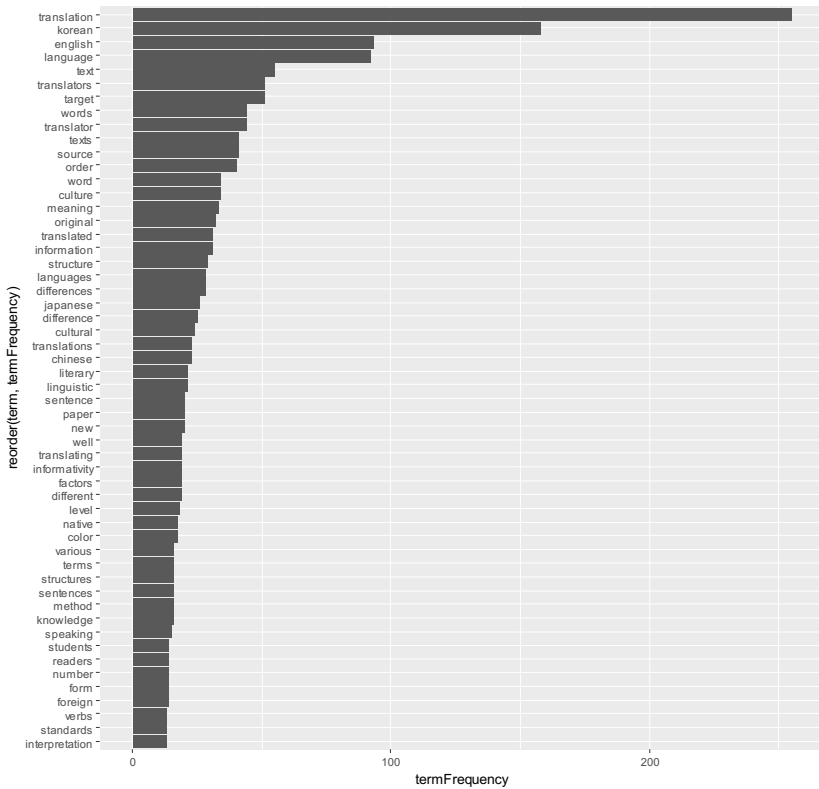
마지막으로 각 분석에서 발견된 주요 키워드를 해석하는 과정에서는 내용 파악이 애매한 키워드의 경우 해당 코퍼스에서 콘코던스 검색을 통해 사용 맥락을 확인하는 절차를 거쳤다.

4. 분석 결과

4.1 발생 빈도 분석

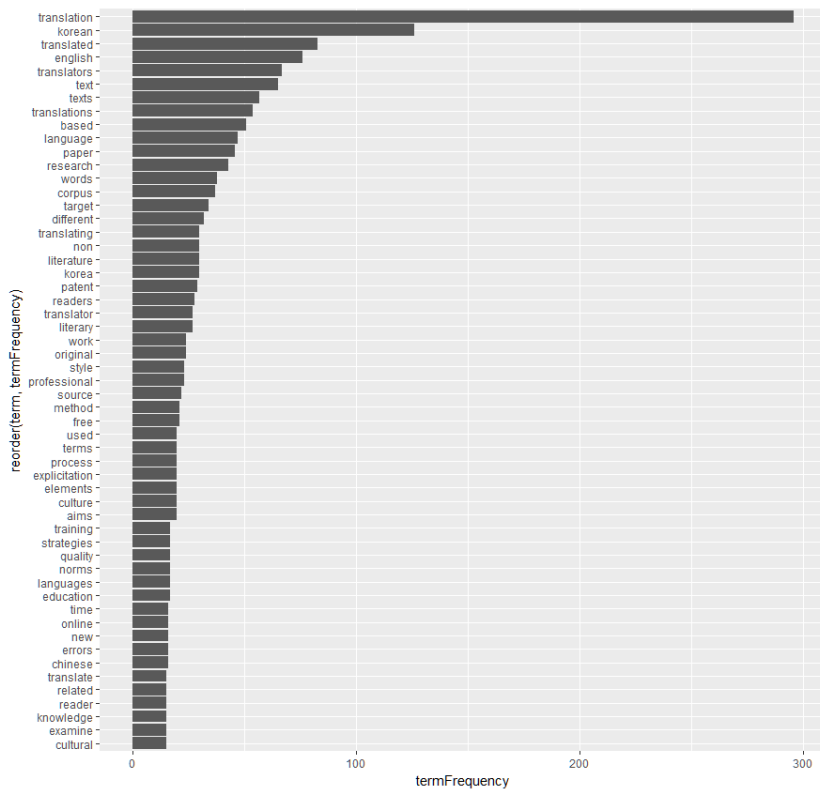
먼저 JTS1과 JTS2에서 어떤 어휘가 주로 나타나는지를 분석해보기로 한다. 이 분석을 위해서 TDM의 문서 행을 JTS1과 JTS2 두 집단으로 나눠 각 단어별 빈도수를 계산하였다. 양 코퍼스의 상위 빈도 어휘는 <그림 2>와 <그림 3>의 그래프에 나와 있다. 그래프는 양 코퍼스의 크기 차이를 고려하여 JTS1은 최소빈도수 13 이상, JTS2는 15 이상인 어휘를 보여준다. 양 코퍼스에서 가장 큰 빈도수를 차지하는 어휘는 translation(번역)인데 번역 연구 논문의 초록에서 이 단어가 가장 많이 나타나는 것은 당연한 결과이다. 그런데 그래프를 보면서 양 코퍼스 간에 상대적 비중이 큰 어휘를 비교하는 것이 쉽지 않다. 테스트 마이닝에서 어휘 빈도수를 시각적으로 쉽게 검토하기 위하여 흔히 사용하는 기법으로 워드클라우드(word cloud)가 있다. 본고에서는 두 코퍼스 간에 상대적으로 많이 발생하는 어휘를 쉽게 알아볼 수 있도록 그림 4와 같은 비교 워드클라우드(comparison word cloud)를 생성하여 어휘 분포를 분석해보기로 한다.

<그림 2> JTS1 상위 빈도 어휘 (최소빈도=13)



<그림 4>의 비교 워드클라우드에는 JTS1과 JTS2 두 코퍼스 안에 포함된 어휘를 상대적 비중에 따라 두 코퍼스 영역으로 분산 배치한 것이다. 좀 더 구체적으로 설명하자면 a라는 단어가 JTS1에서는 21번, JTS2에서는 54번 발생하였다면 평균은 37.5가 된다. 이 평균치를 각 코퍼스의 발생건수에서 빼면 JTS1은 -16.5가 되고 JTS2는 +16.5가 된다. 이 경우 JTS2에서 상대적으로 16.5만큼 더 많이 발생한다는 의미가 되며, 해당 어휘는 그래프에서 JTS2의 영역에 배치되고 글자 크기는 평균과 차이에 비례하여 커진다.

〈그림 3〉 JTS2의 상위 빈도 어휘 (최소빈도=15)



먼저 JTS1에서 특징적으로 나타나는 어휘를 살펴보도록 하자. 가장 눈에 띄는 것은 langague(언어)인데 언어 중에는 English(영어)와 Japanese(일어), chinese(중국어)가 포함되어 있다. 이는 초기에 영어, 일어, 중국어를 분석 언어로 하는 논문이 많았음을 보여준다. 그러나 language란 키워드는 언어구조적 측면을 의미하는 것으로도 해석할 수 있다. 이는 words(단어), subject(주어), verbs(동사), sentence(s)(문장), order(순서), consructions(구조), structures(구조), passive(수동태), surface(표층)와 같은 어휘들과 연관이 있다. 즉, 연구 초기에는 번역문을 언어구조적 측면에서 분석한 논문의 비중이 매우 컸음을 알 수 있다.

또한 source(출발), target(도착), original(원문)과 같은 어휘도 JST1 영역에

회는 텍스트언어학에서 다루는 정보성, 응결성 등의 문제를 언급한 연구의 비중이 높았음을 보여주는데, 초록을 확인해본 결과 이는 특정 연구자의 논문에서 비롯된 것으로 일반적 현상은 아니었다.

이번에는 JTS2 쪽으로 시선을 돌려보면 가장 눈에 띄는 것이 corpus(코퍼스)와 based(기반), corpusbased, cts란 어휘인데, 이는 corpus-base(코퍼스 기반) 연구(CTS)를 지칭하는 어휘로 최근 들어 코퍼스에 기초한 연구가 크게 증가한 상황을 반영한다. 이는 explicitation(명시화), translated(번역된), non-translated(비번역된) 등의 어휘와 긴밀하게 연결되어 있다. explication은 Baker(1993, 1996)가 언급한 번역보편성 중 하나로 번역어와 비번역어 간의 코퍼스 비교 연구를 통해 번역어의 특징을 밝히는 연구의 흐름과 연관이 있다.

discourse(담화), texts(텍스트) 등의 어휘가 JTS2에서 나타나는 것은 번역작품 분석 차원이 JTS1의 단어, 문장 등 언어적 차원에서 텍스트 담화 차원으로 발전한 상황을 반영한다. readers는 번역에서 독자의 역할을 강조하는 최근 연구 동향을 반영하는 것으로 보인다.

번역 이론 면에서 strategies(전략), norms(규범), expectancy(기대) 등의 용어는 Toury(1995)가 선구적 역할을 한 기술적 번역연구와 연관된 어휘로 번역연구 초기와 비교하여 경험적-기술적 연구가 중요한 축으로 자리매김한 것을 보여준다. 또한 functions(기능)란 단어는 원문에 대한 번역문의 충실성에서 번역문 자체의 기능을 중시하는 독일의 기능주의/스코포스 번역학 이론(Nord 1997; Vermeer 1996)이 부각된 상황을 반영한다.

JTS2에서 눈에 띄는 또 다른 특징은 trainee(피훈련자), teaching(교수), learning(학습), educational(교육) 등의 어휘로 최근 들어 통번역 교육에 관한 연구가 크게 늘어났음을 보여준다. 초록을 확인한 결과 online이란 단어도 인터넷을 활용한 통번역교육과 관련된 것으로 나타났다.

그 외에 public은 공공기관 번역, certificates은 자격증, move, subtitles은 영화자막번역, machine은 기계번역, patent, legal은 특허, 법률번역, children은 아동대상번역, avt는 영상매체번역과 연관된 어휘로 최근 번역연구에서 특수번역에 관한 연구가 활성화되고 상황을 반영한다.

이상의 분석결과를 종합하면 국내 번역연구 초기인 2000년대 초에는 번역에서 원문과 번역문 간의 관계, 언어구조적 차이, 바람직한 번역 방식 같은 고

전적인 번역 문제들에 관심을 가졌음을 알 수 있다. 이는 JTS1이 대표하는 시기가 번역학이 국내에 처음 소개되고 본격적인 번역연구가 시작된 시점임을 생각하면 당연해 보인다. JTS2의 어휘를 보면 15년의 시간이 흐르는 동안 국내 번역학 연구는 연구방법, 관심주제, 번역이론 등에서 상당한 발전과 확장이 있었음을 알 수 있다. 그 결과 코퍼스 기반 실증적 연구, 교육 훈련, 전문 번역 등의 연구가 크게 활성화 된 것으로 나타난다.

4.2 상호연관성 네트워크 분석

4.1절에서 시도한 비교 워드클라우드 분석은 JTS1과 JTS2의 나름대로의 특징을 시각적으로 검토하는데 도움이 되지만 각 코퍼스 내의 어휘가 상호 어떤 관계를 가지고 있는지를 직접 보여주지는 않는다. 위에서 분석한 내용은 상당 부분 연구자의 배경지식에 기초한 추론적 해석에 바탕을 둔 것으로 해석이 틀릴 수 있는 여지가 있다. 이를 보완하기 위하여 본 절에서는 어휘 간 상호연관성을 통계적으로 보여주는 네트워크 분석 결과를 검토해보기로 한다. 상호연관성이란 어휘 쌍이 특정 텍스트 단위 내에서 발생하는 빈도를 의미한다. 여기서 단위는 문장, 문단, 텍스트 전체 등 여러 수준에서 정할 수 있는데 본 연구에서는 개별 초록 전체를 분석 단위로 설정하였다. 이는 초록 내에서 같이 발생하는 어휘의 상호 연관성은 논문의 주제 및 내용과 밀접한 관계가 있을 것이라는 전제에 기초한다.

앞서 비교 워드클라우드를 생성하는데 사용한 DTM 데이터가 초록 단위로 만들어진 것이기 때문에 이를 상호연관성 행렬로 변환한 후에 JTS1과 JTS2에서 각각 연관성이 가장 높은 어휘 80개를 추출하여 네트워크 데이터로 변환한 후에 게피란 프로그램을 사용하여 <그림 5>, <그림 6>과 같이 네트워크를 생성하였다. 그림에서 동그라미로 표시된 어휘는 네트워크의 노드(node)라고 하며 노드의 크기는 해당 코퍼스 내에서의 발생빈도에 비례한다. 노드를 연결하는 선은 엣지(edge)라고 하며 엣지의 굵기는 해당 엣지로 연결된 두 노드가 초록 내에서 같이 발생하는 빈도, 즉 상호연관성의 강도를 나타낸다. 또한 노드는 JTS1의 경우는 6개, JTS2의 경우는 7개로 색으로 분류되어 있는데 이를 커뮤니티(community)라 부르며 상호 연관성이 높은 노드의 군집을 의미한다. 본 논

다음에는 시계 반대방향으로 내려가서 보라색으로 표시된 어휘를 보면 error(오류), traditional(전통적), original(원문), japanese(일어), method(방법), translated(번역된) 등이 하나의 군집을 형성한다. 이는 네트워크 중간으로 옮겨 가며 translators(번역사), translation(s)(번역) 등의 중심 노드와 연결되어 있다. 그리고 다시 우상단에 national(국가의), important(중요한), based(기반한), called(불리는), history(역사), chinese(중국), characters(문자), people(사람들), education(교육), students(교육), approach(접근방식), time(시간), foreign(외국의), interpretation(통역), understanding(이해) 등의 어휘들과 연결되어 있다. 이 커뮤니티의 경우는 한 쪽은 일어 번역에서의 오류와 이를 극복하기 위한 번역 방법, 다른 쪽은 중국어 번역의 역사와 한자어의 번역 문제 및 이와 관련된 교육 문제를 다룬 연구의 흐름을 보여준다. 또한 그래프 하단 중앙부에 culture(문화), translating(번역), new(새로운), equivalence(등가) 등의 어휘도 보라색으로 분류되어 있다. 이를 종합적으로 고려하면 특정 외국어 번역의 등가성 문제를 다룬 연구가 JTS1에서 하나의 중요한 축을 형성하고 있는 것으로 판단된다.

다음에는 좌하단에 갈색으로 표시된 커뮤니티를 보면 change(변화), meaning(의미), structure(s)(구조), surface(표층), texts(텍스트), word(단어), order(순서) 등의 어휘로 구성되어 있다. 이는 번역에서 어순, 언어적 구조, 의미 등의 문제가 중요한 연구 주제임을 나타낸다.

우하단의 하늘색으로 표시된 군집은 source(출발), target(도착), text, information(정보), informativity(정보성), knowledge(지식), form(형태), level(층위), standards(기준), differences(차이), literary(문학) 등의 어휘를 담고 있다. 이는 도착어와 출발어 간의 테스트 차원에서의 정보(성) 문제의 차이점을 다룬 연구가 큰 비중을 차지하고 있음을 보여준다. 특히 이들 어휘 중 target, text, source 등의 노드가 네트워크 중앙의 english, korean, language, translation 노드와 비교적 굵은 엣지로 연결되어 있는 것을 볼 때 여기서도 한-영 번역이 주된 연구 대상임을 알 수 있다.

우하단 외곽 쪽으로 빨간색으로 표시된 category(범주), factors(요소), linguistic(언어적), context(맥락) 등의 어휘는 특별히 통일된 주제로 묶기 어려워 보이지만 언어적 범주, 요소 등의 문제를 다룬 연구를 반영한 것으로 여겨진다. 그 외에 그래프 하단 중간에 translator(번역자), cultural(문화적)이란 두 단

어가 파란색으로 표시되어 있는데 어휘 수가 적어 특정 주제를 반영하는 것으로 보기는 힘들지만 그 옆에 있는 culture(문화)와 함께 번역에서 문화가 중요한 주제로 등장하고 있음을 알 수 있다.

이상의 네트워크 분석을 종합하면 JTS1의 경우는 번역을 언어구조적 측면에서 연구한 것이 주류임을 알 수 있다. 이는 앞서 비교 워드클라우드 분석에서 내린 결론과 크게 다르지 않다.

이번에는 <그림 4>로 넘어가서 JST2의 주제적 특징을 살펴보기로 한다. 먼저 래프 상단에 하늘색으로 표시된 커뮤니티를 보면, translating(번역하기), time(시간), terms(용어), functions(기능), style(문체), literary(문학의), work(작품), elements(요소), used(사용된) 등의 어휘가 포함되어 있다. 이는 문학작품에서 문체 번역의 문제를 다룬 연구의 흐름과 번역에서 기능을 중시하는 경향을 보여준다.

시계반대 방향으로 옮겨가서 좌상단에 녹색으로 표시된 군집을 보면, korean(한국어), english(영어), chinese(중국어) 등의 언어를 중심으로 process(과정), cultural(문화적), literature(문학), readers(독자), different(다른) 등의 어휘가 포함되어 있는데, 이는 문학 번역에서 문화적 차원의 번역 문제, 번역 과정과 독자의 역할 등의 문제를 다룬 연구 방향을 보여준다. 그밖에 particular(특정한), present(현재의)는 일반 형용사로 특별한 주제적 의미가 있어 보이지 않으며 public(공공의)은 초록을 검토한 결과 공공기관 번역과 관련 있는 것으로 나타났다.

다음에 그래프 중앙에 갈색으로 표시된 translate(번역하다), meaning(의미), culture(문화), errors(오류), original(원문), discourse(담화), free(자유), translator(s)(번역자) 등의 어휘로 구성된 군집의 경우는 문화적 차이에 따른 번역 방식과 그에 따른 번역 오류 등의 문제를 다룬 연구를 반영한다.

좌하단에 초록색 군집의 경우는 translations(번역물)이란 큰 노드를 제외하면 books(책), titles(제목), words(단어), shifts(변이), examine(검토하다), given(주어진) 등의 어휘로 구성되어 있는데, 영화나 책, 타이틀 번역을 다룬 연구의 비중이 큰 것을 보여준다.

적인) 등의 어휘는 코퍼스 분석이 전문 번역연구에서도 활용되고 있음을 보여 준다. 그 밖의 어휘는 특별한 주제적 의미를 갖고 있는 것으로 보이지 않는다.

마지막으로 이상단에 빨간색으로 표신된 군집은 주제면에서 가장 뚜렷한 집단을 형성하고 있다. training(훈련)이란 중심 노드를 중심으로 learning(학습), teaching(교수) 등의 어휘가 통번역 교육과 관련된 연구가 JTS2에서 중요한 한 축을 형성하고 있음을 나타낸다. online은 온라인 교육이 중요한 approach(접근 방법)로 연구되고 있음을 보여주며, role(역할)은 교육 참여자의 역할도 중요한 연구 주제임을 나타낸다.

이상의 분석을 종합해 보면 JTS2의 경우 문학번역, 코퍼스, 통번역교육, 전문번역 등의 연구가 두드러진 흐름을 형성하고 있다고 결론을 내릴 수 있다. 이는 앞서 비교 워드클라우드 분석의 내용과 크게 다르지 않다.

5. 결론

본 연구에서는 『번역학연구』 학술지 발행 초기와 최근에 발표된 논문의 영문 초록을 분석하여 양 시기 간의 연구 내용의 차이를 분석하였다. 그 결과 15년 정도의 간격을 두고 국내 번역 연구의 흐름이 크게 바뀌었음을 알 수 있었다. 그 흐름의 방향은 (1) 문장, 단어 차원의 언어구조적 연구 → 담화-텍스트적 차원의 연구 (2) 등가적 연구 → 기능적 연구, (3) 정성분석 연구 → 코퍼스를 활용한 실증적 연구, (4) 이론적 연구 → 교육 훈련 등 응용적 연구, (5) 일반번역 연구 → 문학번역, 전문번역 연구 등으로 요약할 수 있다. 이는 국내 번역연구가 내용과 방법 면에서 다양화되고 외연이 크게 확장되었음을 보여준다. 이와 같은 연구 성과에 『번역학연구』를 비롯한 전문학술지가 큰 역할을 했음은 두 말할 나위가 없다.

이상의 연구 결과 외에 본 연구는 전통적인 콘텐츠 분석 기법에서 벗어나 최근 사회학에서 각광을 받고 있는 텍스트 마이닝, 데이터 마이닝 기법을 활용하여 다량의 텍스트를 컴퓨터로 분석하여 근저의 큰 주제적 흐름을 살펴봤다는 데에 그 의미가 있다. 본 연구에서는 비교 워드클라우드와 어휘 상호연관성 네트워크 분석을 시도하였는데, 그 외에도 주제 분석에서 최근 많이 사용되고 있

는 토픽모델링 기법의 활용도 고려해 볼만하다. 본 연구를 위하여 토픽모델링 분석을 실시하기는 하였으나 만족할 만한 수준의 주제 분류가 이뤄지지 않았다. 토픽모델링은 연구자가 토픽의 수를 지정하는 등 다양한 조건을 가지고 실험을 해야 하기 때문에 추가 연구 작업이 필요해 보인다.

연구의 제약점으로는 먼저 학술지 초기 논문인 JTS1의 경우 논문 초록 간에 길이에서 편차가 컸다는 점을 들 수 있다. 따라서 초록의 길이가 긴 논문에서 추출한 키워드의 발생빈도가 상대적으로 높게 나타났을 가능성이 있다. 다만 초록 길이의 편차 영향은 JTS1 내에 국한되며 키워드는 각 코퍼스 내의 주요 연구 주제를 파악하는 용도로만 사용하였기 때문에 JTS2와 비교하여 연구 동향의 변화를 파악하는데는 큰 문제가 되지 않는 것으로 판단하였다. 또한, 본 연구는 『번역학연구』에 실린 논문만을 분석 대상으로 했기 때문에 동 기간에 국내 번역연구의 변화 흐름을 통합적으로 반영하지 않았을 가능성이 있다. 그러나 동 학술지가 번역연구에서 국내 대표적 학술지인 만큼 본 논문의 분석 결과가 번역 연구의 큰 흐름을 보여준다고 해도 무리가 아닐 것이다.

참고문헌

- 김지원 (2000) 「번역학의 발전과 최근 동향에 관한 연구」, 『비평문학』 (14): 115-131.
- 김혜림 (2012) 「중국의 번역연구 동향」, 『번역학연구』 13(4): 29-54.
- 김혜림 (2013) 「한국의 당대(1999-2013) 번역학 연구 동향」, 『번역학연구』 14(5): 95-115.
- 서유경 (2013) 「러시아의 통번역 연구 고찰」, 『통번역학연구』 17(2): 41-68.
- 유효현, 신세인 (2015) 「사회네트워크 분석을 활용한 『한국의학교육』 학회지의 연구 동향 분석」, 『한국의학교육』 27(4): 247-254.
- 이건일, 이종혁 (2015) 「한국 심층학습을 이용한 기계번역 연구 동향」, 『정보과 학회지』 33(10): 40-47.
- 정호정 (2013) 「번역학 전문학술지의 발달과 최근 연구 동향 - 1999년 이후 국내의 대표 학술지 분석」, 『통역과 번역』 15(1): 235-256.

- 조성은 (2014) 「영상번역 연구의 동향과 전망」, 『번역학연구』 15(2): 205-224.
- 최희경 (2015) 「코퍼스를 활용한 번역 연구」, 『번역학연구』 16(2): 317-338.
- 한현희 (2017) 「한국의 통번역 교육 연구 동향에 관한 메타분석 연구」, 『번역학연구』 18(1): 171-202.
- Baker, Mona (1993) 'Corpus Linguistics and Translation Studies: Implications and Applications', in Mona Baker, Gill Francis. & Elena Tognini-Bonell (eds.) *Text and Technology, in Honor of John Sinclair*, Amsterdam: John Benjamins, 233-250
- Baker, Mona (1996) 'Corpus-based Translation Studies: the Challenges that Lie Ahead', in Harold Somers (ed.) *Terminology, LSP and Translation, Studies in Language Engineering in Honour of Iuan C. Sager*, Amsterdam: John Benjamins, 175-186.
- Bailey, Kenneth (1994) *Methods of Social Research*, New York: the Free Press.
- Garten, Yael (2010) *Text Mining of the Scientific Literature to Identify Pharmacogenomic Interaction*, Ph.D Dissertation, Stanford University.
- Nord, Christiane (1997), *Functionalist Approaches Explained*, Manchester: St Jerome.
- Shali, Hai-Jew (2017) *Social Media Data Extraction and Content Analysis*, Hershey, PA.: IGI Global.
- Toury, Gideon (1995) *Descriptive Translation Studies and Beyond*, Amsterdam: Benjamins.
- Vermeer, Hans (1996) *A Skopos Theory of Translation, Some Arguments for and against*, Heidelberg: TEXTconTEXT.
- Wiedemann, Gregor (2016) *Text Mining for Qualitative Data Analysis in the Social Sciences, A Study on Democratic Discourse in Germany*, London: Springer.

[Abstract]

**Past and Present of Korean Translation Research:
A Text Mining-based Analysis
— A Case Study of ‘Journal of Translation Studies’**

Lee, Chang-soo
(Hankuk University of Foreign Studies)

The paper is aimed at shedding light on major changes in the trends of translation research in Korea between the early 2000s and the present. The period of the early 2000s represents the burgeoning of serious academic inquiry into translation issues in Korea, marked by the launch of three major academic journals. The present study employs a text mining approach to tracking down major changes in research trends between the two periods rather than relying on meticulous manual classification according to a pre-defined set of categories as has been attempted by a few earlier studies. The data analyzed in the paper is made of English abstracts compiled from a single journal, ‘the Journal of Translation Studies’ published by the Korean Association of Translation Studies. The data was subjected to word cloud and co-occurrence network analysis. The results show important shifts in research content. They are summarized as (1) from focus on linguistic structure to concern with discourse and text (2) from equivalence-oriented to function-oriented (3) from example/episode-based to corpus-based and empirical, driven by a broad adoption of corpus methodology (4) from theoretical to applied research with a remarkable expansion of training and education and (5) from general to literary and professional translation.

▶ Key Words: translation studies, research trends, text mining, co-occurrence network analysis, Journal of Translation Studies

▶ 주제어: 번역연구, 연구동향, 텍스트 마이닝, 상호연관성 네트워크 분석, 번역학연구

이창수

한국외국어대학교 통역번역대학원 교수

soolee@hanmail.net

관심분야: 코퍼스번역연구, 전산문체학, 디지털 인문학, 비평담화분석, 체계기능 언어학

논문투고일: 2018년 5월 10일

심사완료일: 2018년 5월 30일

게재확정일: 2018년 6월 15일