

## 한국의 중남미 지역연구 네트워크와 중심성 및 무역과 경제에 대한 토픽 변동분석

이재득  
부산대학교 무역학부 교수

## Network, Centrality, and Topic Analysis on Korea's Trade and Economy with Latin America and the Caribbean Area

Chae-Deug Yi<sup>a</sup>

<sup>a</sup>Division of International Trade, Pusan National University, South Korea, South Korea

Received 30 November 2022, Revised 27 December 2022, Accepted 30 December 2022

### Abstract

This study aims to analyze Latin America and the Caribbean papers published in Korea during the past 2000-2020 years. Through this study, it is possible to understand the main subject and direction of research in Korea's Latin America and the Caribbean area. As the research mythologies, this study uses the text mining and Social Network Analysis such as frequency analysis, several centrality analyses, and topic analysis. After analyzing the empirical results, there has been a tendency to change the key words and centrality coefficients between 2000-2010 and 2011-2020 years. During 2011-2020 years, the most frequent keywords were changed from Neoliberalism and culture to policy education, and economy related words. The degree and closeness centrality analyses appeared the higher frequency key words. However, the eigenvector centrality appeared very different from the order of frequency key words. The topic analysis shows that the culture, language, and Neoliberalism were the most important keywords during 2000-2010 years but economy, labor trade, industry, development became the most important keywords during 2011-2020 years in topics.

**Keywords:** Latin America and Caribbean, Text Mining, Network, Frequency, Centrality, Topic Modeling

**JEL Classifications:** F10, C80

<sup>a</sup> First and Corresponding Author, E-mail: [givethaks@pusan.ac.kr](mailto:givethaks@pusan.ac.kr)  
© 2022 The Korea Trade Research Institute. All rights reserved.

## I. 서론

우리나라에 있어 브라질과 아르헨티나, 그리고 멕시코를 포함하는 중남미 지역은 2021년 기준으로 수출은 약 258.2억 달러, 수입액은 약 284.4억 달러를 기록하는 등 많은 교역과 오래된 이민 등의 역사로 볼 때 중요한 지역이다. 그럼에도 불구하고 유럽이나 북미, 아시아 지역국가들에 비해 중남미 전문가들을 제외하고는 상대적으로 일반 국민들의 관심은 다소 미흡한 편이다. 특히 2022년은 멕시코, 아르헨티나, 콜롬비아, 칠레, 코스타리카, 파나마 등 중남미 15개국(이상 수교일자 순)과 수교 60주년을 맞는 해다. 그리고 우리나라는 유사 이래 처음으로 칠레와 2002년 자유무역협정(FTA)을 맺고 2004년부터 자유무역협정이 발효된 이래 경제 및 무역교류를 활발히 확대하고 있다.

그러나 우리나라와 중남미 지역 간의 오랜 이민과 경제 및 무역, 문화교류 등으로 인한 중남미 지역의 중요성으로 비추어 볼 때, 미국과 중국과 일본을 비롯한 아시아와 유럽 등 다 대륙권에 대한 분석연구가 상대적으로 미흡하다. 텍스트 마이닝(Text Mining)은 비정형화 된 텍스트 데이터로부터 키워드를 분석하여 사회적 현상, 추이를 분석한다. 본 연구에서는 주요 키워드를 추출하는 텍스트 마이닝과 사회적 네트워크(Social Network Analysis; SNA)를 통한 네트워크 분석을 통해 경제학적 분석에 주로 사용되는 구조화된 수치자료의 한계를 상호 보완함과 동시에 새로운 분석과 정보를 사용하여 중남미 지역연구에서의 주요 동향과 중요한 이슈를 규명하는데 효율적으로 사용할 수 있다.

그럼에도 불구하고 최근까지 우리나라와 중남미 지역 국가 간의 연구를 분석하는데 있어서 텍스트 마이닝을 사용한 연구키워드에 대한 빈도분석과 중심성 분석(Centrality Analysis) 및 사회적 네트워크 기법을 활용한 우리나라의 중남미 지역연구가 극히 미흡한 편이다. 물론 기존의 중남미 지역에 관한 논문을 분석한 연구들이 있었지만, 대부분 텍스트 마이닝과 SNA를 활용하기 보다는 기존의 문헌연구를 분석하는 기술로 끝나는 감이 있고, 네트워크 중심성과 주제에 대한 분석은 거의 없는 편이다.

최근 2010년대 후반에 와서야 텍스트 마이닝과 SNA를 활용한 중남미 지역연구가 나오고 있는 실정이다. 물론 Im, Doo-Bin and Leem Byung-Hak (2020)과 Choi, Kwon-Jun (2021) 등은 텍스트 마이닝 기법을 사용하여 중남미에 대한 연구를 하였지만, 타 학문분야에 비해 텍스트 마이닝과 SNA 기법을 활용한 지역연구는 극히 미흡하다.

그리하여 본 연구에서 주요 키워드를 추출하는 텍스트 마이닝과 Carter and Feld (2004)과 Poulin, Boily, and Mâsse (2000) 등의 연구에서 처럼 SNA 기법을 통해 계량경제학에서 주로 사용하는 정량분석의 한계를 상호 보완함으로써 중남미 지역연구에서의 주요 동향과 중요한 이슈를 규명하는데 사용한다.

본 연구의 주요 목적은 우리나라의 중남미국가들에 대한 지역연구 논문을 중심으로 중남미 지역연구의 키워드가 무엇인지 살펴보기 위해 2000년대의 중남미 지역에 대한 우리나라의 주요 논문들을 살펴보아서 중남미 연구에서의 중요한 관심사항과 주요 동향을 조사하여 향후 연구의 중요한 관심사항의 심도 깊은 후발 연구를 위한 토대를 제공하는 것이다.

이를 위해 본 연구에서는 비정형 데이터인 텍스트에서 의미와 가치 있는 정보를 찾아내는 기술인 텍스트 마이닝과 단어와 단어 간의 관계를 연결망으로 추출할 수 있게 하는 소셜 네트워크 분석을 주요 분석 방법으로 활용하였으며 토픽(Topic)을 분석하기 위해 머신러닝 기법을 이용하는 Blei, Ng, and Jordan (2003) 등이 고안한 잠재 디리클레 할당 (Latent Dirichlet Allocation, LDA) 기법을 활용하여 분석하였다.

본 연구대상으로는 2000-2020년까지 중남미 지역에 대한 국내 대표적인 중남미 관련 학술지에 게재된 1,195편을 활용하여 분석하였다. 이와 같이 우리나라의 중남미 지역관련 논문을 수집하여, 본 연구의 주요 분석기법은 기존의 지역연구 혹은 특히 중남미지역을 대상으로 거의 사용하지 않았던 분석기법들인 텍스트 마이닝과 SNA, 그리고 중남미 지역연구에서의 키워드와 연구 네트워크와 중심성, 그리고 토픽 등을 분석할 것이다. 그리고 본 연구는 2000-2010년 그리고 2011-2010년으로 10년씩으로 나누

**Table 1.** Korea's Trade(US\$ 1,000, Year 2021)

	Export	Import	Trade Account
Asia	386,133,427	298,333,595	87,799,832
North America	102,616,412	79,583,190	23,033,222
Europe	89,426,982	97,411,551	-7,984,569
EU	47,223,358	66,482,038	-19,258,680
Latin America and Caribbean	25,816,866	28,439,376	-2,622,510
Oceania	15,658,056	35,875,044	-20,216,988
Middle East	15,599,832	65,188,905	-49,589,073
Africa	9,022,390	6,816,379	2,206,011
Rest of the World	126,402	3,445,408	-3,319,006

**Fig. 1.** Korea's Export, Import, and Current Account(2021)



어 빈도수와 중심성, 토픽들이 중남미 지역연구에서 어떤 의미 있는 변동이 있는지 비교분석하고자 한다.

## II. 한국과 중남미 경제협력과 선행연구

### 1. 한국과 중남미 교역동향

기존의 본 절은 왜 중남미 지역에 대한 연구

를 하는 근거와 동기를 보기 위하여 먼저 우리나라와 중남미 지역 국가들 간의 교역동향을 살펴본다. 2021년 한 해 동안 우리나라와 중남미 간의 무역액을 살펴보면, 수출액은 <Table 1>과 <Fig. 1>에서 나와 있듯이, 대륙권별로는 무역액은 아시아와 북미 그리고 유럽이 상위를 차지하고 있지만 대양주, 중동, 그리고 아프리카 국가들보다는 무역액이 큰 것으로 나타났다.

특히 우리나라는 중남미 국가들에 수출은 약 258.2억 달러, 수입액은 약 284.4억 달러를 기록하였고, 그리하여 무역수지는 26.2억 달러의

**Table 2.** Annual Trade between Korea and Latin America and Caribbean(US\$ 1,000, %)

Year	Export	% Change	Import	% Change	Trade Account
2000	9,368,955	8.4	3,262,608	13.9	6,106,347
2001	9,729,633	3.8	3,445,010	5.6	6,284,623
2002	8,863,550	-8.9	3,742,716	8.6	5,120,834
2003	8,801,566	-0.7	4,594,485	22.8	4,207,081
2004	11,562,623	31.4	6,651,155	44.8	4,911,468
2005	14,986,882	29.6	7,016,806	5.5	7,970,076
2006	20,590,704	37.4	9,732,234	38.7	10,858,469
2007	25,781,307	25.2	11,323,924	16.4	14,457,383
2008	33,267,324	29	13,755,559	21.5	19,511,765
2009	26,763,576	-19.5	11,647,795	-15.3	15,115,781
2010	36,187,294	35.2	14,644,745	25.7	21,542,549
2011	40,131,402	10.9	20,165,274	37.7	19,966,128
2012	36,746,769	-8.4	19,722,605	-2.2	17,024,164
2013	36,327,984	-1.1	18,363,537	-6.9	17,964,447
2014	35,897,793	-1.2	18,294,433	-0.4	17,603,360
2015	30,676,988	-14.5	15,970,810	-12.7	14,706,178
2016	25,440,755	-17.1	15,217,121	-4.7	10,223,634
2017	28,095,242	10.4	17,080,590	12.2	11,014,652
2018	27,767,562	-1.2	19,331,755	13.2	8,435,807
2019	26,337,382	-5.2	20,341,719	5.2	5,995,662
2020	19,501,465	-26	20,381,062	0.2	-879,597
2021	25,816,866	32.4	28,439,376	39.5	-2,622,510

적자를 기록하고 있다. 이러한 우리나라의 중남미 지역국가들에 대한 무역수지 적자는 우리나라가 수출하는 공산품보다 중남미로부터 더 많은 광물자원과 농산물 등을 수입하고 있기 때문이다.

한편, 다음 <Table 2>는 한국과 중남미 지역 간 2000년부터 2021년까지의 연도별 무역동향을 나타내고 있다. 그리하여 2000년 우리나라의 중남미 지역에 대한 수출은 약 93.7억 달러를 기록하였고 수입은 약 32.6억 달러를 기록하였으며, 우리나라의 대 중남미 지역에 대한 무역수지는 약 61.1억을 기록하였다.

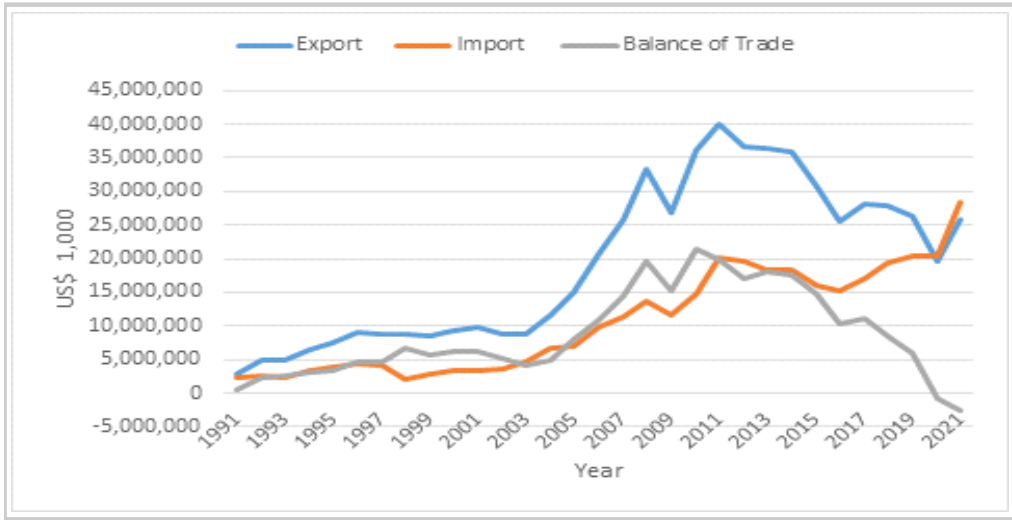
그러나 우리나라의 중남미 지역에 대한 수출액은 2004년 115.6억 달러, 수입은 66.5억 달러를 기록하였고, 무역수지는 49.1억 달러를 각각

기록하는 등 큰 폭으로 증가하기 시작하였다. 그리하여 2006년에는 우리나라의 수출액이 205.9억 달러를 기록하였고, 2008년에는 936.2억 달러를 기록하는 등 수출이 크게 증가하였고, 수입액도 2007년 113.2억 달러, 2011년 201.7억 달러를 기록하는 등 교역액이 증가하였다.

무역수지를 보면 1991년 이래 우리나라는 최근 2019년까지 무역수지 흑자를 기록하였는데 2010년에 215.4억 달러를 기록하였고 그 후 2012년 등 2010년대 초반 세계 경제위기로 인해 교역액이 감소하여 무역수지도 조금씩 감소 추세를 보이고 있다.

특히 2019년부터 2022년 현재 Covid-19 팬데믹으로 인하여 세계적으로 교역액 크게 감소

**Fig. 2.** Annual Trend of Korea's Export, Import, and Current Account with Latin America and Caribbean



하였고 우리나라도 2020년 중남미 지역에 대한 수출액이 전년도보다 26.0%로 대폭 감소하였고 중남미 지역으로 부터의 수입은 0.2% 증가하여 2021년 처음으로 중남미 지역에 대해 무역적자가 8.8 억 달러가 발생하였다.

그러나 2021년에는 우리나라의 수출은 2021년보다 32.4%가 증가한 258.2억 달러, 수입은 39.5%가 증가한 284.4억 달러를 기록하여, 우리나라의 중남미 지역에 대한 무역수지는 262.3억 달러의 커다란 적자를 기록하는 등, 우리나라와 중남미 지역 국가 간 무역은 세계경제 불황과 Covid-19 팬데믹 등으로 인한 세계 경기에 대해 민감하게 반응하고 있는 것으로 나타났다. 이러한 한국과 중남미 지역 간 수출액, 수출입, 그리고 무역수지 무역동향은 (Fig. 2)에서도 잘 나타나고 있다.

## 2. 텍스트 마이닝과 네트워크 기법에 의한 선행 연구

선행연구를 보면, 먼저 텍스트 마이닝을 활용한 선행연구로는 경영경제학과 관광, 사회학과 신문방송학, 문헌언론정보학, 그리고 마케팅

등의 다양한 사회과학 분야에서 사용되고 있으나, 국제지역학 분야에서는 연구가 미흡한 편이다. 하물며 중남미 지역에 대한 텍스트 마이닝과 소셜 네트워크 기법에 의한 연구는 극히 미흡하다.

그리하여 텍스트 마이닝과 관련된 국제지역 연구 외 타분야의 기존 문헌연구들을 보면, Ko, Hyun-Jeong (2021)은 공급사슬에 영향을 미치는 요인을 2015부터 2021까지 발표된 280편의 논문을 추출하여 상위 10개의 키워드를 분석하고 있다. Ko, Min-Hwan and Yun-Mi Park (2022)는 빅데이터 기반의 텍스트 마이닝을 활용한 포스트 코로나 시대의 미래관광에 대한 연구를 하였다. Suh, Jeong-Meen (2020)은 텍스트 마이닝 접근법을 통하여 APEC에서의 지역무역협정(RTA)과 자유무역협정(FTA)의 수립과 발산에 대한 연구를 하였다.

한편, 토픽분석은 별로 없지만, Lee, Jee-Hoon and Jung-Suk Kim (2000)은 2002부터 2019까지 국내 무역학 주요 학술지인 무역학회지, 국제통상연구, Journal of Korea Trade에 수록된 논문을 이용하여 토픽을 분석하였고, Yi, Chae-Deug (2022)은 Blei, Ng, and Jordan

(2003) 연구를 바탕으로 LDA 알고리즘을 활용하여 국제무역연구에 대한 데이터 마이닝과 네트워크 및 중심성과 토픽분석을 하였다.

중남미 지역에 관한 텍스트 마이닝과 소셜 네트워크를 활용한 연구는 거의 없는 실정이다. Cha, Kyung-Mi (2011)와 Hong, Uk-Heon (2012)은 각각 한국의 라틴아메리카 지역연구 동향에 대한 문헌연구를 하였다. 그러나 이들 연구들은 텍스트 마이닝과 소셜 네트워크 분석을 활용하지 않고 있다. Lee, Sung-Jun and Yoon, Taek-Dong (2014)은 한국과 브라질 간 국가 문화 차이에 관한 연구를 하였지만, 역시 텍스트 마이닝을 활용한 연구라기보다는 역시 기존의 문헌을 종합하여 기술하고 있다. Kim, Seung-Nyeon and Lee Sang-Jik (2016)은 한국의 중남미 ODA와 경제협력과 관련성 분석을 하였지만, 주로 상관분석과 인과관계 분석을 하였다.

Im, Doo-Bin and Leem Byung-Hak (2020)은 국내 포르투갈어권과 중남미 지역연구동향을 KCI 자료를 활용하여 단어 빈도분석과 구조적 등위성을 중심으로 텍스트 마이닝 분석을 활용한 분석하였지만, 토픽에 대한 연구가 없다. Choi, Kwon-Jun (2021) 중남미 전문 학술지들을 중심으로 한국의 중남미 문학연구 동향에 대한 연구를 하였다. Yi, Chae-Deug (2002)은 우리나라의 국제무역연구동향을 소셜 네트워크와 토픽모델링 기법을 통해 분석하였다.

이와 같이 우리나라와 중남미 지역 간의 오랜 이민과 경제 및 무역, 문화교류 등으로 인한 중남미 지역의 중요성으로 비추어 볼 때, 북미와 중국과 동아시아, 그리고 유럽 등의 타 대륙권에 대한 무역과 경제협력 연구에 비해 상대적으로 미약하다. 더구나 텍스트 마이닝 기법을 이용한 키워드의 빈도분석 그리고 사회적 네트워크를 이용한 중심성 분석과 토픽 모델링 기법을 활용한 연구가 거의 없는 실정이다.

### Ⅲ. 연구분석의 이론적 배경

#### 1. 텍스트 마이닝(Text Mining)과 빈도분석

Kim, Nam-Gyu, Lee Domg-Hoon, Choi Ho-Chang, and Wong William Xiu Shun (2017)의 연구에 나타나 있듯이, 텍스트는 인류가 정보를 표현하고 전달하는 데에 사용해 온 가장 대표적인 수단이다. Paranyushkin (2011)에 의하면, 텍스트 마이닝의 분석방법으로는 텍스트 분류, 텍스트 군집화, 정보추출, 네트워크분석, 감성분석 등이 있는데 이들은 텍스트에 나타난 주요 개념을 추출하는 것은 물론 다른 개념과의 관계를 파악하는데 유용하다. 그리하여 본 연구는 이러한 텍스트 마이닝 분석 방법을 통해 국내 중남미 지역의 논문들을 기반으로 실증분석 이전에 전처리를 진행하고, 텍스트에서 출현하는 키워드의 빈도 분석과 연결성을 통해 키워드 간의 연관관계를 도출하여, 이를 기반으로 관계 동향을 분석한다.

#### 2. 소셜 네트워크와 중심성 분석

소셜 네트워크 분석은 Kwahk, Kee-Young (2017)에 의하면, 분석 대상 및 분석 대상들 간의 사회적 관계를 이용하여 연결망 구조를 체계적으로 분석할 수 있다. 이러한 네트워크 분석의 구조적 특성을 이해하기 위해 중심성(centrality), 집중성(centralization), 그리고 밀도(density) 등의 지표가 사용되는데, Park Jun-Suk, Kim Chang-Sik Kim, and Kwahk Kee-Young (2016)과 Yi Chae-Deug (2022) 등이 이러한 기법을 사용하여 분석하고 있다. 그리하여 본 연구는 이러한 텍스트 마이닝과 소셜 네트워크의 중심성 기법을 통해 우리나라의 중남미 지역에 대한 논문들을 기반으로 실증분석을 한다. 실증분석 이전에 전처리를 진행하고, 텍스트에서 출현하는 키워드의 빈도 분석과 연결성을 통해 키워드 간의 연관관계를 도출하여, 이를 기반으로 관계 동향을 분석한다.

본 연구에서는 이를 위해 네트워크에서 중요

한 키워드로 이루어진 어떤 노드 또는 링크가 다른 노드들에 비해 더 중심적임을 나타내는 중심성 분석은 연결 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality), 아이겐벡터 중심성(Eigenvector)과 에고 연결(Ego Degree) 등을 구하여 분석하였다<sup>1)</sup>.

(1) 연결(degree) 중심성:

중심성 지표로서 가장 많이 사용되는 것 중의 하나는 연결(Degree) 중심성이다. 연결 중심성은 네트워크에서 노드로 표시된 액터(actor)에 직접적으로 연결된 링크의 수를 말한다. 특정 노드의 연결정도는 네트워크 내에서 액터의 파워에 대한 가장 간단하고 효과적인 지표로 사용되기 때문에 노드 자체의 중심성을 나타내는 지표로 활용된다. 본고에서는 네트워크의 연결정도는 네트워크 내 임의의 노드에 연결된 총 연결선의 수를 연결정도(degree)로 나타내는데 Kwahk Kee-Young (2016)과 Yi Chae-Deug (2022)의 연구를 참조하면 다음과 같이 간단히 요약하여 정의한다.

먼저  $g$ 개의 노드로 표시되는 액터(actor)를 가지고 있는 네트워크에서 액터  $i$ 의 연결 중심성은 ( $C_D(N_i)$ ) 다음과 같은 식으로 나타낸다.

$$C_D(N_i) = \sum_{j=1}^g x_{ij}, i \neq j.$$

여기서  $g$ 는 액터의 개수,  $\sum_{i=1}^g x_{ij}$ 는 액터  $i$ 가 ( $g-1$ )개의 다른 액터와 갖는 연결관계의 계수로  $x_{ij}$ 는 연결되어 있지 않으면 0 또는 연결되어 있으면 1의 값을 갖는다.

(2) 근접(Closeness) 중심성

근접(Closeness) 중심은 연결 중심성과는 달리 네트워크 내에서의 간접적 연결을 또한 고려하여 액터와 다른 액터 간의 거리를 강조하

는 중심성 개념이다. 그리하여 근접 중심성은 해당 네트워크의 전체적 관점에서 직접효과뿐만 아니라 간접효과도 포함하는 노드의 중심성을 고려하는 개념이다. 따라서 근접 중심성 [ $C_C(N_i)$ ]은  $\sum_{j=1}^g d(N_i, N_j)$ 는 액터  $i$ 와  $j$  간의 최단 경로거리의 합이며,  $g$ 는 액터의 개수를 나타낸다고 하면 다음과 같이 정의된다.

$$C_C(N_i) = \frac{1}{\sum_{j=1}^g d(N_i, N_j)}, i \neq j.$$

(3) 매개(Betweenness) 중심성

매개(Betweenness) 중심성은 네트워크상에서 임의의 액터 혹은 노드가 직접 연결되어 있지 않는 액터 혹은 노드들 간에 매개자를 통해 매개하는 정도를 나타내는 것으로 Freeman (1979)은 매개 중심성( $C_B(N_i)$ )은 다음과 같이 정의하고 있다.

$$C_B(N_i) = \sum_{j < k} \frac{g_{jk}(N_i)}{g_{jk}}, i \neq j \neq k$$

여기서  $g_{jk}$ 는 두 액터  $j$ 와  $k$  간의 최단경로 개수,  $g_{jk}(N_i)$ 는 두 액터  $j$ 와  $k$  간의 최단경로 가운데 액터  $i$ 를 포함하고 있는 경로의 개수를 말한다.

(4) 아이겐벡터(Eigenvector) 중심성

아이겐벡터(Eigenvector) 중심성은 임의의 노드와 연결되어 있는 이웃 노드가 얼마나 중요한가를 함께 고려하는 것으로 Bonacich (2007)가 주장하듯이, 연결 관계뿐만 아니라 한 노드와 연결된 다른 노드의 중요성을 가중치를 고려한 중심성 지표로서 액터  $i$ 의 아이겐벡터의 중심성( $C_E(N_i)$ )은 다음과 같이 정의한다.

$$C_E(N_i) = \lambda \sum_{j=1}^g x_{ij} C_E(N_j), i \neq j.$$

여기서  $\lambda$ 는 아이겐 값을 나타내고  $x_{ij}$ 는 액터  $i$ 와  $j$  간의 0과 1로 구성된 연결관계의 값을 나타내는 값이다. 따라서 강한 아이겐벡터 중심성을 갖고 있는 노드는 그 네트워크상에서

1) Kwahk (2019), Yi (2022)참조.

강한 연결 중심성을 갖고 있다.

### (5) 에고 연결(Ego Degree)과 응집성

Poulin, Boily, and Mâsse (2000)과 Carter and Feld (2004), 그리고 Knoke and Yang (2008)등에 의하면, 어떤 특정 노드를 중심으로 네트워크의 최적화 개념은, 밀도(Density)로서 표시되는데 밀도란 한 네트워크에서 노드사이의 연결이 이루어진 정도를 나타내 것으로 응집성 개념이다.

## 3. 토픽 모델링

토픽(Topic)은 머신러닝 기반의 하나로 단어 나 문서의 집합에서 더 큰 주제로 묶어낼 수 있는 분석기법이다. 그리하여 토픽 모델링(Topic Modeling) 중에서 사용하는 알고리즘은 주로 Blei, Ng, and Jordan (2003)이 창안한 LDA 모형을 가장 많이 사용한다. LDA는 주어진 연구 문헌들의 요약이나 키워드 등 관찰된 변수를 통해 전체 문헌집합의 주제와 각 연구문헌별 주제비율, 각 단어들이 각 주제에 포함될 확률들을 계산해 낸다.

그리하여 각 문서에 어떤 주제들이 존재하는 지에 대해 확률모형으로 주제별 단어분포나 문서별 주제의 분포를 추정할 수 있다. LDA 기법에 의한 토픽분석은 주제별 단어 분포를 바탕으로 주어진 문서에서 발견된 단어의 확률분포를 분석함으로써, 앞으로 어떤 주제를 다루고 있을지를 예측한다.

Blei, Ng, and Jordan (2003)의 LDA 기법에 의한 토픽분석은 다음과 같이 구해진다. 먼저 아래 식에서 M은 전체 논문의 개수, w는 단어를 나타낸다고 할 때, N은 특정 문서에 속한 단어의 개수를 나타내며 포아송(Poisson) 분포를 한다. Z는 해당 단어가 속하는 주제를 나타낸다.  $\alpha$ 와  $\beta$ 는 초모수(hyper parameter)이며, 말뭉치는 M 개의 문서의 수집에 의해 정해진다.  $\theta$ 는 각 논문에 해당되는 토픽분포로서 모두 d 개의 디리클레(Dirichlet) 분포를 하고 있다고 상정한다. 그리하여 초모수인  $\alpha$ 와  $\beta$ 가 주어졌다고 가정할 때, 토픽  $\theta$ , Z, 그리고 w의 결합

분포는 다음과 같이 주어진다.

$$p(\theta, Z, w|\alpha, \beta) =$$

$$p(\theta|\alpha) \prod_{n=1}^N p(Z_n|\theta)p(w_n|Z_n, \beta).$$

먼저  $\theta$ 에 대해 적분하고 Z에 대해 합하면 다음과 같은 한계(Marginal) 분포를 얻을 수 있다.

$$p(w|\alpha, \beta) =$$

$$\int p(\theta|\alpha) \left( \prod_{n=1}^N \sum_{Z_n} p(Z_n|\theta)p(w_n|Z_n, \beta) \right) d\theta.$$

다음 각각의 논문들의 한계 확률의 곱을 취하면, 다음과 같은 말뭉치의 확률을 얻을 수 있다.

$$p(D|\alpha, \beta)$$

$$= \prod_{d=1}^M \int p(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{Z_{dn}} p(Z_{dn}|\theta_d)p(w_n|Z_{dn}, \beta) \right) d\theta_d.$$

그리하여 LDA 분석을 통해 사전에  $\alpha$ 와  $\beta$ , 토픽의 개수를 정한 후 Z,  $\theta$ , d를 구한다. 그리고 LDA는 논문 내의 단어들, 즉 w를 관측해 나가면서 각 단어마다 임의의 주제 Z 값을 정한다. 그리고 이 결과에 따라  $\theta$ ,  $\varphi$ 의 디리클레분포를 업데이트하고 반복과정을 통해 가장 가능성이 높은 Z값을 찾아내고,  $\theta$ , 그리고 d를 추정하기 위해 반복적인 시뮬레이션을 통해 값을 얻는다. 그리고 특정 주제에 관한 논문에는 그 주제에 관해 자주 등장하는 단어들을 균집화하여 잠재된 주제를 추론한 후, 주제별 단어 분포를 바탕으로 주어진 문서에서 발견된 단어의 확률 분포를 분석한다.



## IV. 실증분석 결과

### 1. 연구분석 자료와 전처리 방법

본 연구는 우리나라에서 2000년부터 2020년까지 발간된 중남미 지역연구 관련 학회 연구 논문 1,195편을 한국 연구재단의 KCI 문헌 자료들을 이용하여 분석하였다. 본 연구에 활용된 해당 학술논문지는 라틴아메리카 연구 556편, 중남미연구 455편, 그리고 포르투갈-브라질 연구 184편이며 총 대상논문은 1,195편으로 나타나 있다. 그리고 본 연구에서는 해당년도 보다는 10년별로 나누어 2000-2010 기간과 2011-2020년 기간으로 두 개의 시기로 나누어 이 두 시기에 있어 중남미 지역연구의 초록에 나와 있는 키워드와 네트워크의 중심성, 그리고 토픽에 어떤 변동이 있는지 비교분석하였다.

본 연구분석에 앞서 전처리를 보면, 실증분석에 필요한 모든 논문 자료는 한글 키워드는 없는 경우가 많았기 때문에 통일을 기하기 위하여 영문 키워드를 수집하였다. 그리고 본 연구에서는 영문 키워드에서는 명사만을 대상으로 하였고, 대명사, 형용사, 부사, 동사, 전치사, 접속사, 그리고 관사 등은 제외하였다. 수집된 키워드는 연구자에 의해 단어 정규화, 불용어 제거 등 데이터 전처리 작업을 진행하였다. 데이터 전처리 과정은 동일한 의미를 지닌 단어를 하나로 통합하였으며, 연구논문의 키워드에서 주로 쓰는 paper, research, 그리고 study 등과 별 의미가 없고 본 연구의 분석과 연관이 없는 낱말은 제거하였다.

그 다음 본 연구분석을 위해서는 Netminer 4를 사용하였다. 그리하여 키워드의 출현을 기반으로 하는 빈도분석과 네트워크 분석(Network Analysis)을 위해 여러 가지 중심성(Centrality) 분석을 중심으로 하였으며, 토픽(Topic) 모델링을 통해 토픽분석을 하였다.

### 2. 중남미 연구에 대한 빈도분석

텍스트 마이닝에 의한 중남미 지역국가들에 대한 정치, 경제, 사회 문화 분야 등에 나타난

전체 키워드들의 빈도분석 결과들을 <Table 3>에서 보면, 2000-2010년 동안 중남미지역에서 연구 키워드는 Neoliberalism이 18회로 가장 높게 나타났고, 그 다음에는 Culture 14 회, Mexico 10회, Corea 8회, language 8회, 그리고 myth와 democratization 등이 각각 7회로 높은 순으로 나타났다.

그리고 2011-2020년 동안에는 중남미지역 연구들에서 키워드는 policy 20회, education, 20회, economy 16회, development 15회, Brasil 15회, language 13회, labor 12회, literature 10회, 그리고 identity 10회 등의 순으로 높게 나타났다.

그러나 우리나라와 중남미 지역국가들에 대한 경제와 무역에 연관된 연구들을 살펴보면, 2000-2010년 동안에는 중남미국가들에 대한 경제와 무역에 대한 연구가 비교적 낮은 편이었지만, 2011-2020년 사이에는 economy 16회, development 15회, labor 12회 등으로 비교적 높게 나타났다.

그리고 경제와 연관된 migration과 industry와 inequality 등도 각각 6회로 나타났다. 그리하여 우리나라의 중남미국가들에 대한 연구는 2000년대 초기에는 정치, 문화, 그리고 사회에 대한 연구보다 경제와 무역에 관한 관심이 높은 편이 아니었지만, 2010년대 들어서는 무역과 경제와 연관된 관심이 높은 것으로 변해간 것을 알 수 있다.

그리하여 2000-2010년 동안 중남미지역은 중남미에서 신자유주의와 중남미 문화에 등 정치와 사회와 문화 등에 관련된 연구가 가장 높게 나타났고, 멕시코와 한국과 중남미에 관련된 연구가 많이 나타난 것을 알 수 있다.

그러나 2011-2020년으로 넘어가면서 멕시코 보다는 브라질에 대한 연구가 더 많이 나타났고, 2010년대 후반으로 가면서 좀 더 세부적이고 소프트한 정책과 교육, 그리고 정치보다는 경제무역 및 산업, 노동 등에 관한 연구가 더 많이 나타나 우리나라의 중남미 지역국가들에 대한 연구는 2010년대에 있어서 경제와 개발 등에 대한 관심이 더 높은 것으로 나타나, 중남미 지역국가들에 대한 경제와 무역에 대한 연구가 비교적 높게 나타나고 있다.

**Table 3.** Frequency of Key Word in Latin America and Caribbean

2000-2010	Frequency	2011-2020	Frequency
Neoliberalism	18	policy	20
culture	14	education	20
Mexico	10	economy	16
Corea	8	development	15
language	8	Brasil	15
myth	7	language	13
democratization	7	labor	12
Politics	7	literature	10
Latin America	6	identity	10
literatura	6	history	8
Brazil	6	power	8
government	6	migration	8
politics	6	Portuguese	8
immigration	5	culture	7
humanos	5	crisis	7
Cervantes	5	wind	6
globalization	4	movement	6
Venezuela	4	law	6
maquiladora	4	inequality	6
Ingua	4	industry	6
entre	4	cooperation	6
teaching	4	Borges	6

### 3. 네트워크 중심성 분석

네트워크 분석은 Netminer 4를 통해 분석하였으며, 본 연구에서 네트워크 분석은 우선 키워드 출현 빈도가 높은 대략 상위 50-60개 단어를 추출하여 각 키워드에 대한 네트워크 분석을 위해 중심성 분석을 하였다. 네트워크 분석은 먼저 텍스트 파싱 및 텍스트 필터링 작업을 수행한 후의 데이터를 기반으로 연관분석을 수행하였다. 네트워크 분석 또한 텍스트 마이닝에 의한 빈도분석과 마찬가지로 2000년-2010년, 2011년-2020년으로 두 기간으로 구분하여 비교분석하였다.

#### (1) 중남미 지역연구의 2000-2010년 중심성 분석

먼저 2000년에서 2011년까지 네트워크 중심성 분석 결과, <Table 4>를 보면, 연결 중심성의 경우 cultura(0.08), democratization(0.08), dilemma(0.08), estrangeira(0.07), government(0.07), 그리고 language(0.07) 등의 순으로 높게 나타났다. 간접효과를 고려한 근접 중심성의 경우는 연결 중심성 순위와 조금 다르게 나타났다. 근접 중심성은 cultura(0.14), democratization(0.10), dilemma(0.09), estrangeira(0.14), government(0.09), 그리고 language(0.16), environment(0.13), immigration(0.14), strategy(0.13) 등으로 나타나 language 순위가 제일 높게 나타났다.

**Table 4.** 2000–2010 Centrality of Latin America and Caribbean Study Network

Key Word	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality	Density
cultura	0.076923	0.140412	0.031927	0.000075	0.5
democratization	0.076923	0.101935	0.144231	0.282482	0.571429
dilemma	0.076923	0.091293	0.025485	0.386951	0.571429
culture	0.067308	0.147474	0.201176	0.000083	0.52381
estrangreira	0.067308	0.139235	0.012323	0.000096	0.619048
government	0.067308	0.09096	0.001027	0.315194	0.761905
language	0.067308	0.155284	0.307506	0.000072	0.47619
maquiladora	0.067308	0.100903	0.086072	0.000379	0.333333
party	0.067308	0.091126	0.012229	0.442917	0.761905
regime	0.067308	0.091293	0.037497	0.353946	0.714286
Teaching	0.057692	0.138847	0.009055	0.000061	0.666667
export	0.057692	0.099892	0.026139	0.000439	0.466667
hispanoamericana	0.057692	0.106737	0.020164	0.000981	0.466667
labor	0.057692	0.090794	0	0.431508	1
literatura	0.057692	0.121873	0.144978	0.000839	0.4
Ingua	0.057692	0.139625	0.018484	0.000079	0.733333
neoliberalism	0.057692	0.090794	0	0.353725	1
amor	0.048077	0.106509	0.013256	0.000554	0.6
economy	0.048077	0.100093	0.038648	0.000403	0.6
investment	0.048077	0.112014	0.167289	0.000411	0.6
mujeres	0.048077	0.107427	0.062547	0.000672	0.6
myth	0.048077	0.118399	0.026326	0.00006	0.3
coreana	0.038462	0.118399	0.026326	0.000018	0.166667
entre	0.038462	0.084058	0.026326	0.000037	0.166667
environment	0.038462	0.122774	0.174571	0.000147	0.5
fantastica	0.038462	0.106282	0	0.001017	1
immigration	0.038462	0.135821	0.194175	0.000383	0.333333
strategy	0.038462	0.133636	0.205377	0.000071	0.5
Neoliberalism	0.028846	0.081581	0.025734	0.0096	0
Brasileira	0.028846	0.117562	0	0.000046	1
capital	0.028846	0.08917	0.013256	0.000057	0.333333
construction	0.028846	0.11784	0	0.000065	1
mexicano	0.028846	0.09317	0.00028	0.000203	0.333333
para	0.028846	0.094585	0.050784	0.000175	0.333333
policy	0.028846	0.123688	0.167289	0.00229	0.333333
politics	0.028846	0.081448	0.012914	0.126123	0.333333
reform	0.028846	0.119535	0.026139	0.000373	0.333333
study	0.028846	0.129135	0.004014	0.000079	0.333333

한편, 매개 중심성의 경우는 democratization (0.14), language(0.31), literatura(0.14), investment (0.17), environment(0.17), immigration(0.19), strategy(0.20), 그리고 policy(0.16) 등으로 language와 strategy 등이 높게 나타났다. 키워드의 중요성을 고려하는 아이젠벡터 중심성의 경우는 역시 연결, 근접, 매개 중심성의 순위와는 많이 다르게 나타났는데, dilemma(0.39), government(0.32), party(0.44), regime(0.35), labor(0.43), 그리고 neoliberalism(0.35) 등이 높게 나타났다.

2000년에서 2010년까지 응집성을 나타내는 밀도의 순위는 역시 앞의 연결 중심성이나 근접 및 매개 중심성과는 다르게 순위가 나타났다. cultura(0.50), democratization(0.57), dilemma (0.57), estrangeira(0.62), government(0.76), 그리고 language(0.48) 등으로 높게 나왔다. party와 regime도 0.70보다 더 높게 나타났고, labor, neoliberalism 그리고 Brasileira 등도 밀도가 아주 높게 나타났다.

그러나 경제와 무역에 대한 키워드의 연결 중심성은 멕시코의 수출가공지역인 maquiladora (0.06), export(0.06), labor(0.06), economy(0.05), 그리고 investment(0.05) 등으로 중위권 이하로 정치경제사회문화 등의 분야와 비교해볼 때, 비교적 낮게 나타났으며, 근접중심성과 연결중심성보다는 조금 높게 나타났으며, 경제와 무역과 연관된 연구의 근접중심성의 순위는 비슷하게 나타났다. 매개 중심성의 경우에는 investment(0.17)만 높게 나왔으나 대부분 경제 (0.04)와 수출(0.03) 등으로 비교적 낮게 나타났다.

그리고 아이젠벡터 중심성의 경우도 labor를 제외하고는 수출과 경제, 그리고 maquiladora 프로그램도 아주 낮게 나타나. 2000-2010 기간 동안에는 경제보다는 정치와 사회, 그리고 문화에 대한 연구가 상대적으로 많이 나타나고 있다. 그리고 응집성을 나타내는 밀도의 경우는 export(0.47), economy(0.06), investment (0.06)로 각각 비교적 높지 않게 나타났다.

## (2) 중남미 지역연구의 2011-2020년 중심성 분석

2011년에서 2020년까지 중남미지역 네트워크 중심성 분석 결과, <Table 5>를 보면, 연결 중심성의 경우 policy(0.31), education(0.22), language(0.18), cooperation(0.14), power(0.13), labor(0.12), school(0.12), inequality(0.10), literature(0.10), system(0.10), development (0.09), economy(0.09), 그리고 resource(0.09) 등의 순으로 높게 나타났다. 그리하여 2000년에서 2011년 동안 높은 순위를 차지하였던 것과는 달리 정책과 교육, 언어, 노동, 불균등, 경제와 산업개발 등이 높은 순으로 나타났다.

2011년에서 2020년 동안 중남미지역에서는 간접효과를 고려한 근접 중심성의 경우에는 연결 중심성 순위와 대체로 비슷하게 나타났으나 조금 더 근접 중심성 계수가 더 높게 나타났다. 그리하여 근접 중심성 순위가 높았던 policy(0.47), education(0.43), language(0.37), cooperation(0.38), power(0.35), labor(0.35), school(0.39), inequality(0.34), literature(0.34), system(0.36) 등으로 근접 중심성도 높게 나타났다.

2011년에서 2020년 동안 중남미지역에서 매개 중심성의 경우는 policy(0.30), education(0.14) 순으로 높게 나타났으나, 그 외 language(0.10), cooperation(0.05), power(0.05), labor(0.04), school(0.04) 등으로 나타났다. inequality(0.08), literature(0.09), 그리고 system(0.39) 등으로 나타났다.

그러나 아이젠벡터 중심성의 경우는 역시 연결, 근접, 매개 중심성의 순위와는 많이 다르게 나타났는데, cooperation(0.23), power(0.41), capacity(0.44), resource(0.33), factor(0.47), 그리고 wind(0.43), 그리고 production(0.20) 등이 높게 나타났다. 그리하여 2011-2020년 동안에는 중남미 지역의 경제와 직결되는 생산요소, 능력 등과 그리고 정치적인 정파, 힘 등에 대한 관심이 높아져 그와 연관된 연구가 많이 나타났다.

마지막으로 2011년에서 2020년 동안 중남미 지역의 응집성을 나타내는 밀도의 순위는 연결

**Table 5.** 2011–2020 Centrality of Latin America and Caribbean Study Network

Key Word	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality	Density
policy	0.307692	0.473166	0.296469	0.067117	0.126812
education	0.217949	0.432897	0.14347	0.024755	0.198529
language	0.179487	0.374469	0.09751	0.008146	0.241758
cooperation	0.141026	0.38389	0.050541	0.230567	0.381818
power	0.128205	0.354875	0.045963	0.407665	0.377778
labor	0.115385	0.352823	0.04097	0.007643	0.305556
school	0.115385	0.386319	0.042905	0.009319	0.388889
inequality	0.102564	0.340997	0.075286	0.024743	0.214286
literature	0.102564	0.340997	0.092156	0.001567	0.142857
system	0.102564	0.363324	0.089636	0.031158	0.25
Portuguese	0.089744	0.367702	0.038814	0.012975	0.380952
achievement	0.089744	0.38389	0.01293	0.009645	0.571429
capacity	0.089744	0.337229	0.008215	0.444207	0.666667
development	0.089744	0.361174	0.028685	0.056926	0.238095
economy	0.089744	0.352823	0.051949	0.005423	0.190476
resource	0.089744	0.308275	0.005498	0.326346	0.619048
state	0.089744	0.372186	0.0197	0.105269	0.380952
analysis	0.076923	0.321255	0.017657	0.005157	0.333333
crisis	0.076923	0.331731	0.025262	0.003411	0.333333
factor	0.076923	0.286566	0.000844	0.469907	0.8
industry	0.076923	0.340997	0.010547	0.00986	0.266667
migration	0.076923	0.337229	0.006433	0.008826	0.6
reform	0.076923	0.35695	0.027867	0.005726	0.466667
satisfaction	0.076923	0.322955	0.003386	0.005832	0.6
word	0.076923	0.322955	0.01467	0.002877	0.4
Language	0.064103	0.296303	0.002824	0.003039	0.6
community	0.064103	0.346809	0.024976	0.004048	0.4
identity	0.064103	0.311421	0.021767	0.000986	0.3
law	0.064103	0.289282	0.007761	0.005246	0.4

중심성 순위에서 높게 나타났던 policy(0.13), education(0.20), language(0.25) 보다 cooperation(0.38), power(0.38), labor(0.31), school(0.39) 등이 높게 나타났다. 그리고 연결 중심성 순위에서 이들 보다 낮게 나타났던 Portuguese(0.38), achievement (0.57), capacity(0.67) 등으로 높게 나타난 경우가 많이 나타나고 있다. 그리하여 capacity, achievement, cooperation, power, labor 등이

높게 나타나 역시 중남미 지역은 2010년대 중요시 하였던 경제와 발전에 관련된 연구가 많이 나타나고 있다.

그러나 2011-2020년 동안의 경제와 무역에 대한 키워드의 중심성을 살펴보면, 경제무역에 대한 키워드들의 중심성이 2000-2010동안보다 다소 높게 나타나있다. 먼저 연결중심성은 labor(0.12), inequality(0.10), economy(0.09), development(0.09), resource (0.09),

**Table 6.** Five Topics by LDA in MCMC

	1 <sup>st</sup> Keyword	2 <sup>nd</sup> Keyword	3 <sup>rd</sup> Keyword	4 <sup>th</sup> Keyword	5 <sup>th</sup> Keyword
Topic 1	culture	Coreana	Neoliberalism	myth	democratization
Topic 2	Neoliberalism	government	immigration	democratization	Teaching
Topic 3	Neoliberalism	culture	Coreanas	government	coreanos
Topic 4	Neoliberalism	myth	barataria	Language	Cervantes
Topic 5	Neoliberalism	entre	Coreana	cultura	myth

industry(0.09), 그리고 migration(0.08) 등으로 비교적 높게 나타나 있다. 근접중심성의 경우 도 연결중심성과 비슷한 순위로 나타났지만, 비교적 높은 순위로 나타났다.

간접적인 효과를 포함하는 매개중심성도 연결중심성보다 다소 낮게 나타났으며 경제와 무역과 연관된 연구의 순위는 비슷하게 나타났다. 그리고 아이겐벡터 중심성의 경우는 inequality(0.02), development(0.03), economy(0.05), resource(0.01), industry(0.01), 그리고 migration(0.01) 등으로 나타나서 경제와 개발에 대한 중심성은 비교적 높게 나타났다. 응집성을 나타내는 밀도의 경우 labor(0.31), development(0.24), economy(0.19), resource(0.62), industry(0.27), 그리고 migration(0.60) 등으로 나타나 대체로 아이겐 벡터의 중심성보다 상대적으로 다소 높게 나타나 있다.

그리하여 2000-2010년 동안보다 2011-2020년 동안의 중남미 국가들에 대한 연구동향을 중심으로 비교해보면, 정치, 정치, 문화보다는 경제와 무역, 그리고 노동, 산업과 개발 등에 관한 중요성과 관심이 높아지고 있는 것을 알 수 있다. 이것은 중남미 지역국가들에 대한 연구 키워드가 상대적으로 볼 때, 더 많이 나타나, 경제와 무역 및 산업에 대한 연구에 대한 숫자가 증가하고 있는 변동이 나타난 것을 알 수 있다.

## 5. 토픽(Topic) 분석

본 장에서는 중남미 지역연구 논문의 텍스트 본문을 키워드 중심으로 숨겨진 의미구조를 발견하는 토픽분석을 통해 이들 논문을 어떤 키

워드 묶을 수 있는지 찾아낸다. 그 후 토픽을 구성하고 있는 키워드 분포를 바탕으로 연구본문들에서 발견된 키워드들의 분포를 분석하여 추출해낸 키워드들이 어떤 토픽을 다루고 있을지를 예측한다.

LDA는 논문별 토픽별 키워드 분포의 결합분포에 의해 연구논문에서 사용된 키워드를 추출하고, 그 연구논문을 나타내는 토픽들이 어떤 단어들로 구성되어 있는지 그 확률분포를 구해야 한다. 본 연구에서는 토픽의 수를 5개로 선정하여 MCMC(Monte-Carlo Markov Chain) 방법을 활용하여 시뮬레이션을 반복하여 각 키워드가 할당받은 확률을 추정하였다.

### (1) 2000-2010 중남미 지역연구 토픽분석

그리하여 먼저 아래와 같이 키워드 빈도수 상위 약 35개를 기준으로 토픽에서 발생할 확률이 높은 키워드를 구하였다. 그 다음에 <Table 6>과 <Table 7>에서 나타나 있듯이, 다섯 개의 토픽들과 확률이 높은 키워드를 토픽별로 다섯 개씩 구하였다. 그리하여 먼저 2000-2010년 동안 다섯 개 키워드가 <Table 6>에 나타나 있다.

<Table 7>에서는 다섯 개의 개별적인 토픽별로 앞에서 구한 빈도분석에서 나타난 상위 순위 키워드 50개 중에서 출현 확률을 나타내고 있는데 그 확률이 가장 높은 키워드 다섯 개를 추출한 것이 <Table 6>에 나타나 있는 것이다. 즉 <Table 7>은 특정한 무역연구 토픽에 특정 무역연구 키워드가 나타날 확률을 보여주고 있다.

**Table 7.** 2000–2010 Probability of Key Word in 5 Topics Latin America and Caribbean Study

Key Word	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
cultura	0.01837	0.00931	0.01177	0.00883	0.02315
democratization	0.02167	0.02361	0.00502	0.00532	0.00360
dilemma	0.00496	0.00955	0.01786	0.01347	0.00538
culture	0.05732	0.00728	0.03886	0.01710	0.00775
estraneira	0.00532	0.00747	0.00620	0.00685	0.00892
government	0.00728	0.03495	0.02222	0.00896	0.00972
language	0.00757	0.00991	0.01095	0.01077	0.00815
maquiladora	0.01396	0.01777	0.01534	0.00831	0.00657
party	0.00578	0.00798	0.00776	0.00505	0.00700
regime	0.01171	0.00650	0.00776	0.01253	0.02036
Teaching	0.01002	0.02091	0.01002	0.01138	0.00784
export	0.00652	0.00822	0.00598	0.00325	0.00516
hispanoamericana	0.00673	0.01006	0.00634	0.01095	0.00656
labor	0.00623	0.00940	0.00850	0.00504	0.01027
literatura	0.02155	0.01613	0.00466	0.00572	0.00564
Ingua	0.00945	0.01160	0.01587	0.01040	0.01633
amor	0.00977	0.01256	0.00480	0.01364	0.00993
economy	0.00741	0.00718	0.00749	0.00826	0.00493
investment	0.00865	0.00708	0.00815	0.00787	0.00631
mujeres	0.00705	0.01036	0.00901	0.00591	0.00340
myth	0.02863	0.01947	0.01585	0.02449	0.02238
Coreana	0.03922	0.01639	0.01393	0.01012	0.02630
entre	0.00844	0.00728	0.01118	0.00964	0.03188
environment	0.00911	0.00541	0.00785	0.00856	0.00564
fantastica	0.00675	0.00914	0.00737	0.00670	0.00719
immigration	0.00527	0.03454	0.01451	0.01713	0.01278
strategy	0.01133	0.00505	0.00856	0.00766	0.00766
Neoliberalism	0.03345	0.04977	0.06041	0.07524	0.06707
brasileira	0.00547	0.00867	0.00737	0.00646	0.00694
capital	0.00786	0.00650	0.00941	0.01015	0.01064
construction	0.01042	0.00528	0.00609	0.00813	0.01045
mexicano	0.01194	0.01372	0.01318	0.01096	0.00919
policy	0.00729	0.01360	0.00688	0.01221	0.01116
politics	0.00700	0.00384	0.01023	0.01015	0.00508
Hispanics	0.00575	0.00884	0.01456	0.00682	0.00608
Language	0.00487	0.01356	0.01844	0.02329	0.01095

**Table 8.** Five Topics by LDA in MCMC

	1 <sup>st</sup> Keyword	2 <sup>nd</sup> Keyword	3 <sup>rd</sup> Keyword	4 <sup>th</sup> Keyword	5 <sup>th</sup> Keyword
Topic 1	Economy	policy	culture	crisis	Brasil
Topic 2	Language	Borges	industry	crisis	policy
Topic 3	policy	labor	wind	power	Brasil
Topic 4	education	language	Portuguese	community	identity
Topic 5	policy	labor	literature	Portuguese	migration

그리하여 <Table 7>에서 토픽1에서 키워드 culture가 나타날 확률이 0.057을 가지고 있고, 그 다음 나타날 확률이 높은 키워드는 Coreana 0.039, Neoliberalism 0.033, myth 0.029, democratization 0.022 등의 순으로 높게 나타나고 있다. 그리하여 토픽 1의 레이블(Label)은 '문화와 한국 및 신자유주의 및 민주화'로 정할 수 있다.

토픽 2에서는 키워드 Neoliberalism 0.050, government 0.035, immigration 0.035, democratization 0.024, 그리고 Teaching 0.021 등의 순으로 높은 확률을 가지고 있으며, 이는 '신자유주의와 정부와 이민' 등으로 규정한다. 토픽 3에서는 키워드 역시 Neoliberalism 0.060, culture, Coreanas 0.025, government 0.022, 그리고 Coreanos 0.021 등의 순으로 확률이 높게 나타나, '신자유주의와 문화와 한국'으로 토픽의 레이블을 정한다.

토픽 4에서는 역시 키워드 Neoliberalism 0.075, myth 0.025, Barataria(바라타리아 베이 지역) 0.024, Language 0.023, 그리고 Cervantes(세르반테스 문화보급 기관) 0.022 등의 순으로 높게 나타나 역시 '신자유주의와 신화 및 언어' 등으로 나타난다. 토픽 5에서는 키워드 Neoliberalism 0.067, entre(between us) 0.032, Coreana 0.026, cultura 0.023, 그리고 myth 0.022 등의 순으로 높게 나타나서, '신자유주의와 소통 및 문화'로 토픽 5의 레이블로 정한다.

따라서 2000-2010년 동안 중남미 지역의 토픽분석 결과, Culture, Neoliberalism, Coreana 그리고 Language 등이 중요한 키워드로 나타나 신자유주의와 문화와 한류 등이 5개 토픽들

에 있어 중요한 영향을 미치고 있는 것으로 나타났다.

그리하여, <Table 6>에 나타나 있듯이, 2000-2010년 동안 중남미 지역에 대한 연구를 무역과 경제 측면에서 살펴보면, 5개의 토픽 가운데 정치와 문화 등에 대한 키워드인 culture와 Neoimperialism 등이 주요 키워드로 나타나고 있으나, 한국(Corean)에 대한 키워드는 나타나고 있으나 무역과 경제와 연관된 키워드가 전혀 나타나고 있지 않다.

이것은 2000-2010년 동안에는 한국에서 중남미 지역에 대한 정치, 사회, 그리고 문화 등에 대한 연구와 관심이 높았지만, 한국이 중남미 지역국가들에 대해 경제와 무역에 대한 관심과 연구가 상대적으로 낮았다는 것을 의미한다.

따라서 <Table 7>에 나타나 있듯이, 경제와 무역에 연관된 키워드 들인 maquiladora, export, labor, economy, 그리고 investment 등이 추출한 5개의 토픽에서 나타나 모두 다 확률이 상대적으로 낮게 나타나서 5개의 키워드에 속하지 못하는 것으로 나타났다.

## (2) 2011-2020 중남미 지역연구 토픽분석

2011-2020년 동안 다섯 개 키워드는 <Table 8>에 나타나 있으며, <Table 9>에서 5개의 개별적인 토픽별로 앞에서 구한 빈도분석에서 나타난 상위 순위 키워드 약 35개 중에서 출현 확률이 가장 높은 키워드 다섯 개를 추출한 것이 <Table 8>에 나타나 있다.

즉 <Table 8>에서 토픽1에서 키워드 Economy가 나타날 확률이 0.066을 가지고 있



고, 그 다음 나타날 확률이 높은 키워드는 policy 0.059, culture 0.041, crisis 0.033, 그리고 Brasil 0.028등의 순으로 높게 나타나고 있다. 그리하여 토픽 1의 레이블(Label)은 '경제와 정치 그리고 문화 브라질'로 규정할 수 있다.

토픽 2에서는 Language 0.038, Borges(깊이를 알 수 없는 상상력과 지식, 의미가 짙게 농축된 글과 문학), 0.035, industry, crisis 0.032, 그리고 policy 0.025 등의 순으로 높은 확률을 가지고 있으며, 이는 '언어와 산업 정책' 등으로 규정한다. 토픽 3에서는 policy 0.110, labor 0.061, wind 0.040, power 0.038, 그리고 Brasil 0.029 등의 순으로 높게 나타나고 있다. 그리하여 '정책과 노동 및 정파와 브라질'로 토픽의 레이블을 정한다.

토픽 4에서는 키워드 education 0.206, language 0.043, Portuguese 0.026, community 0.023, 그리고 identity 0.022 등의 순으로 높게 나타나 교육과 포르투갈어 및 공동체" 등으로 나타난다. 마지막으로 토픽 5에서는 키워드 policy 0.114, labor 0.057, literature 0.050, Portuguese 0.023, 그리고 migration 0.021 등의 순으로 높게 나타나서, '정책과 노동 및 포르투갈과 이민'으로 토픽 5의 레이블로 정한다.

따라서 위의 분석결과 2011-2020년 동안의 중남미 지역의 토픽분석 결과는 2000-2011년 동안의 중남미 지역의 토픽분석 결과와 상당히 다르게 나타났다. 즉 2011-2020년 동안 중남미 지역의 토픽은 문화와 한류 등 보다는 경제와 정책, 포르투갈과 브라질 등에 대한 관심이 많이 일어나 이에 토픽들이 2011-2020년 동안 토픽들에 있어 중요한 요소가 되고 있는 것으로 나타났다.

그러나, <Table 9>에 나타나 있듯이, 2011-2020년 동안의 중남미 지역연구를 무역과 경제 측면의 키워드와 토픽을 살펴보면, 그 전의 2000-2010년 동안의 토픽과 키워드가 비교적 많이 변동된 것을 알 수 있다. 즉 2000-2010년 동안에는 문화에 대한 키워드인 culture와 특히 정치적인 Neoliberalism 등이 주요 키워드로 나타난 반면, 2011-2020년 동안의 토픽에서 그 전의 10년 동안에는 전혀 나타나지 않았던

경제와 무역과 연관된 키워드들인 economy, industry, labor, 그리고 migration 등이 나타나고 있다.

멕시코의 미국에 대한 수출의 전초기지로 활용하였던 maquiladora 프로그램이 NAFTA로 대체되면서 수출가공구역이 사라졌지만, 특히 경제부문의 주요 키워드인 economy는 중요한 첫 번째 토픽에서 첫 번째 키워드로 나타나서, 2011-2020년 동안에 있어 2010년대에는 중남미 지역국가들에 있어서도 경제이슈가 아주 중요한 관심사이고 그에 대한 연구가 많은 것으로 나타난 것을 알 수 있다.

그리하여 <Table 9>에 나타나 있듯이, 경제와 무역에 연관된 키워드 들인 economy가 첫 번째 토픽에서 첫 번째 키워드로 나타날 확률이 0.066로 가장 높게 나타난 것을 비롯하여, 두 번째 토픽에서 industry, 그리고 세 번째 토픽에서 labor, 그리고 다섯 번째 토픽의 migration 등이 추출한 5개의 토픽에서 나타나 비교적 높은 확률로 나타나 5개의 키워드에 속하는 것으로 나타나 있다.

그리하여 2000-2010년 동안의 연구대상 기간 보다 그 후의 10년 동안인 2011-2020년 동안의 경우에서 한국의 중남미 지역연구에서 경제와 무역에 대한 관심과 중요성이 많이 나타난 것을 알 수 있다. 그리하여 한국의 중남미 지역국가들에 대한 연구의 중요성이 정치, 문화, 사회적인 측면보다 상대적으로 비교적 2010년을 넘어가면서 최근 10년 동안 연구의 초점과 관심이 경제와 무역에 연관된 토픽이 상대적으로 높아져 가고 있는 것으로 나타났다.

## V. 결론

본 연구는 2000년부터 2020년까지 중남미 지역 연관 학회지들에 게재된 1,195편을 대상으로 테이트 마이닝과 소셜 네트워크 기법과 토픽을 중심으로 분석하였다. 그리고 2000년대인 2000-2010년 그리고 2010년대인 2011-2020년 두 기간으로 나누어 중남미 지역에 있어 어떤 변동이 있는지 비교 분석하였다. 그리하여 본 연구의 결과를 요약하면 다음과 같다.

**Table 9.** 2011–2020 Probability of Key Word in 5 Topics Latin America and Caribbean Study

Key Word	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
policy	0.05927	0.02474	0.10988	0.01754	0.11434
education	0.00841	0.00798	0.00568	0.20626	0.00911
cooperation	0.00976	0.00469	0.02568	0.00923	0.01898
power	0.00883	0.01734	0.03788	0.00506	0.01072
labor	0.01243	0.00919	0.06067	0.00814	0.05681
school	0.01156	0.00905	0.00954	0.00990	0.01407
inequality	0.00930	0.01367	0.00497	0.01273	0.01233
literature	0.02508	0.02344	0.00422	0.00485	0.04958
system	0.02437	0.01693	0.02111	0.00571	0.00664
Portuguese	0.00764	0.00688	0.00777	0.02648	0.02328
resource	0.01572	0.00657	0.01135	0.01457	0.00890
state	0.00489	0.01324	0.00788	0.01294	0.01552
analysis	0.00833	0.01230	0.01456	0.01294	0.01370
crisis	0.03318	0.03234	0.01421	0.00626	0.01581
factor	0.00917	0.00838	0.01770	0.00517	0.01280
industry	0.00292	0.03315	0.01083	0.00706	0.01742
migration	0.01587	0.01687	0.01611	0.00698	0.02139
word	0.01470	0.00906	0.00860	0.00701	0.00433
Language	0.01119	0.03852	0.01241	0.01779	0.00462
community	0.00769	0.01337	0.00674	0.02331	0.00849
identity	0.00807	0.01727	0.02219	0.02173	0.01408
wind	0.00821	0.01515	0.03999	0.01330	0.00461
History	0.00634	0.01004	0.00781	0.00434	0.00581
culture	0.04096	0.02253	0.01714	0.01398	0.01904
institution	0.00619	0.01114	0.00885	0.00878	0.00795
pollution	0.01046	0.00347	0.01225	0.00604	0.01347
production	0.00608	0.01136	0.00617	0.00497	0.00913
government	0.00891	0.00571	0.00602	0.01062	0.01271
history	0.00682	0.02102	0.01485	0.02040	0.01576
crime	0.01735	0.01367	0.02808	0.01649	0.00478
discrimination	0.00606	0.00402	0.00920	0.00719	0.00717
drug	0.01055	0.00473	0.00917	0.00709	0.00652
politics	0.00997	0.00808	0.00263	0.01145	0.00737
quota	0.02172	0.00774	0.00308	0.00243	0.01703
Borges	0.02632	0.03577	0.00591	0.00956	0.01999
Brasil	0.02751	0.01906	0.02851	0.01563	0.00987
Development	0.00989	0.01663	0.01735	0.01338	0.00633
literatura	0.01856	0.00464	0.00566	0.01192	0.00433
Economy	0.06640	0.00230	0.00303	0.00221	0.01418

첫째, 우리나라 중남미 지역 논문을 대상으로 텍스트 마이닝 기법을 이용하여 논문들의 키워드의 빈도수를 추출해본 결과, 2000-2010년 동안은 신자유주의와 문화에 등과 관련된 연구가 많이 나타난 반면, 2011-2020년 동안은 소프트한 정책과 교육, 그리고 경제 및 산업 등에 관한 연구가 더 많이 나타났다. 둘째, 중남미 지역 논문을 대상으로 논문들을 대상으로 2000-2010년 동안의 네트워크 중심성 분석을 한 결과 먼저, 연결 중심성 경우는 cultura, democratization, dilemma, estrangeira, government, 그리고 language 등의 순으로 나타났다지만, 간접효과를 고려한 근접 중심성의 경우는 language, cultura, estrangeira, 그리고 immigration 등으로 순으로 다르게 나타났다. 매개중심성의 순위는 language가 제일 높게 나타났다고 그 다음에는 strategy 가 높게 나타났다. 2011-2020년 동안은 연결 중심성의 경우에는 2000-2010년 동안과는 달리 정책과 교육, 언어, 노동, 불균등, 경제와 산업개발 등이 높은 순으로 나타났다.

셋째, 2000-2010년 동안 중남미지역에서 아이젠벡터 중심성의 경우는 연결, 근접, 매개 중심성의 순위와는 많이 다르게 나타났는데, party, labor, neoliberalism, 그리고 regime 등의 순으로 높게 나타났다. 2011-2020년 동안은 아이젠벡터 중심성의 경우 역시 연결, 근접, 매개 중심성의 순위와 다르게 나타났는데, factor, capacity, wind, power 등의 순으로 높게 나타났다.

넷째, 2000-2010년 동안 중남미 지역의 예고 연결성 및 예고 밀도로 분석한 결과, 2011년-2020년 동안 중남미지역의 밀도의 순위는 연결 중심성 순위에서 높게 나타났던 policy, education, language 보다 capacity, achievement, cooperation, power, labor, school, Portuguese 등이 더 높게 나타나 경제와 발전에 관련된 연구가 많이 나타났다.

다섯째, LDA 모형을 사용하여 5개 토픽모델링 분석을 한 결과, 2000-2010년 동안에는 한류문화와 신자유주의 관련된 토픽이 많았던 반면, 2011년-2020년 동안에는 경제와 정책, 포르투갈과 브라질 등에 대한 토픽이 많이 나타나,

중남미 지역연구 토픽이 한류문화와 신자유주의에서 중남미 지역의 경제정책과 개발로 바뀌고 있음을 알 수 있다.

그리고 마지막으로 한국의 중남미 지역국가들과의 연구키워드와 중심성 분석에서 모두 그전의 2000-2010년에 비해, 그 후 2011-2020년 동안 연구의 관심과 중요성이 정치, 사회, 문화적인 측면에서 경제와 무역으로 상대적으로 변화가 있는 것으로 나타났다. 이러한 경향은 특히 토픽모델링 분석에서 추출한 5개의 토픽에서 economy, industry, labor, 그리고 migration 등이 상대적으로 높은 출현 확률로 나타나 있는 주요한 5개의 키워드 안에 모두 포함되어 있는 것으로 확인 되었다.

따라서 본 연구에서 추정 결과, 2000-2010년 동안에는 문화와 언어 그리고 신자유주의 등에 연관된 연구가 많았지만, 2011년-2020년 동안에는 중남미 지역의 경제정책과 산업과 개발에 관한 연구가 많이 나타났다. 그러나 2010년대를 기점으로 중남미 지역에 대한 연구의 관심이 정치, 사회, 문화적인 것에서 상대적으로 경제와 무역과 관련된 것으로 변하고 있는 것으로 나타났다.

중남미 지역 분야에서 소셜 네트워크를 이용한 기존 연구가 거의 없는 실정이지만, 본 연구는 중남미 지역을 대상으로 데이터 마이닝과 소셜 네트워크 중심성 분석, LDA 기법 등에 의한 토픽분석 등을 연구하여 최근 20년간 중남미 지역 분야의 연구주제와 연구동향을 두 기간으로 나누어 분석하고 있다. 그리하여 본 연구결과는 2000-2010년보다 2011-2020년 동안에는 더욱더 중남미 국가들은 무역과 경제에 대한 관심이 높아지고 있는 것으로 나타났다. 그리하여 우리나라의 무역기업들도 이러한 것을 감안하고 좀 더 무역을 확대하기 위하여 경제적인 면에서 중남미 지역국가들과 경제 및 무역협력을 확대하고 중남미 시장특성에 맞는 적극적인 수출전략으로 나아갈 필요가 있다. 그리고 2000-2010년대 동안 중남미 연구에서 중심적인 역할을 하였던 정치와 언어와 문화 외에도, 2011-2020년 동안은 경제와 무역 분야가 중심적인 이슈로 나타났으므로 정치, 경제, 사회, 문화 분야의 연구자 간 협력연구를 증대

할 필요가 있어 보인다.

그러나 본 연구는 좀 많은 국내외 학회의 중남미 연구자료를 수집하여 연도별로 분석한다면 좀 더 엄밀한 분석과 해석을 할 수 있을 것이다. 그럼에도 불구하고 우리나라 중남미 지역 연구 분야에서 텍스트 마이닝과 소셜 네트

워크, 토픽분석에 의한 연구가 거의 없다. 그리하여 본 연구는 이들의 기법을 응용해서 우리나라의 중남미 지역 연구의 네트워크와 토픽에 있어 2010년을 기점으로 정치, 사회, 문화적인 측면에서 경제와 무역으로 관심과 중요성의 증가로 연구동향이 바뀌고 있는 것을 분석하였다.

## References

- Blei, D. M., A. Y. Ng and M. I. Jordan (2003), "Latent Dirichlet Allocation", *Journal of machine Learning research*, 3, 993-1022.
- Bonacich, P. (2007), "Some unique properties of eigenvector centrality," *Social Networks*, 29(4), 555-564.
- Carter, W. C. and S. L. Feld (2004), "Principles relating social regard to size and density of personal networks, with applications to stigma," *Social Networks*, 26(4), 323-329.
- Cha, Kyung-Mi (2011), "Latin American Regional Study Trend and Individual Nation Study", *Cross-Cultural Studies*, 22, 203-221.
- Choi, Kwon-Jun (2021), "A Study on the Research Trends of Latin American Literature in Korea -Centering on Expert Academic Journals", *Latin America and Caribbean Studies*, 40(3), 199-244.
- Freeman, L.C. (1978), "Centrality in social networks conceptual clarification", *Social Networks* 1, 215-239.
- Han, Jin-Sung and Ji-Hwan Yoon (2016), "Activation Strategies of the 20th BIFF using Social Big Data Text Mining Analysis", *Journal of Tourism Sciences*, 40(1), 133-145.
- Hong, Uk-Heon (2012), "Issues and Trends in Latin American Studies in Korea", *Asia Review*, 2(2), 149-1609.
- Im, Doo-Bin and Leem Byung-Hak (2020), "Analysis of Research Trends in Lusophone Studies in Korea using Text Mining", 39(4), *Latin America and Caribbean Studies*, 179-214.
- Kim, Nam-Gyu, Dong-Hoon Lee, Ho-Chang Choi, and William Xiu Shun Wong (2017), "Investigations on Techniques and Applications of Text Analytics", *The Journal of Korean Institute of Communications and Information Sciences*, 42(2), 471-492.
- Kim, Seung-Nyeon and Lee Sang-Jik (2016), "On the Relationship between ODA and Economic Cooperation among Korea and Latin American Countries", 35(1), *Latin America and Caribbean Studies*, 40(3), 277-305.
- Knoke, D. and S. Yang (2008), *Social Network Analysis*, SAGE, Thousand Oaks.
- Ko, Hyun-Jeong (2021), "Analysis on the Research Topics of Supply Chain Risk Management using Text Mining", *E-Trade Review*, 19(3), 65-83.
- Ko, Min-Hwan and Yun-Mi Park (2022), "A study on future tourism in the post-COVID-19 era using text mining based on big data: Focused on LAN cable tours", *International journal of Tourism and Hospitality*. 36(5), 79-92.
- Kwahk, Kee-Young (2017), *Social Network Analysis*, the 2<sup>nd</sup> edition, Cheongram.
- Lee, Jee-Hoon and Jung-Suk Kim (2000), "A Study on the Research Trends in Int'l Trade Using Topic modeling", *Korea Trade Review*, 45(3), 55-69.

- Lee, Sung-Jun and Yoon, Taek-Dong (2014), “ The National Cultural Differences Between Korea and Brazil - A Critical Review of Five Cross-National Comparison Researches”, *Journal of International Area Studies*, 18(2), 27-54.
- Paranyushkin Dmitry (2011), Identifying the Pathways for Meaning Circulation using Text Network Analysis, *Nodus Labs*, 26, 1-26.
- Park, Jun-Suk, Chang-Sik Kim, and Kee-Young Kwahk (2016), “Investigation of Research Trend in Hotel Domain Using Text Mining and Social Network Analysis”, *Journal of Tourism and Leisure Research*, 28(9), 209-226.
- Poulin, R., M. C. Boily, and B. R. Masse (2000), “Dynamical systems to define centrality in social networks”, *Social networks*, 22(3), 187-220.
- Suh, Jeong-Meen (2020), Measuring Convergences and Divergences in APEC RTAs/FTAs: a text-mining approach, *APEC Study Series*, 20-02, KIEP.
- Yi, Chae-Deug (2022), “A Big Data Analysis on Research Keywords, Centrality, and Topics of International Trade using the Text Mining and Social Network”, *Korea Trade Review*, 47(4), 137-159.