

A Review of Ethnic Minority Tourism Research Using LDA Topic Modeling and Network Analysis: With a Focus on International Tourism Journals

LDA 토픽모델링과 네트워크 분석을 통한 소수 민족관광 연구동향 분석: 국제 관광학 저널 중심으로

Yuan Lee¹

이유안¹

¹ Adjunct Professor, Dept. of Future Convergence, Silla University, Korea, totitre@naver.com

Abstract: This study examines research trends in ethnic tourism published between 2000 and the first half of 2025 in SSCI-indexed tourism journals. Articles were identified through Web of Science, Google Scholar, and the official websites of SSCI-indexed tourism journals, and 155 studies were selected based on established screening criteria. Employing network analysis with UCINET and Latent Dirichlet Allocation (LDA) topic modeling in Python, the study identified key structural patterns. Network centrality analysis, based on degree, eigenvector, betweenness, and closeness measures, revealed that tourism development, community and culture consistently occupied central positions, serving as conceptual hubs in ethnic tourism research. Other significant keywords included tourism impact, tourism business, stakeholder group, ethnic minority, and ethnic village, reflecting research themes at the intersections of development, community, culture, and identity. The LDA analysis produced five dominant topics: (1) heritage & poverty reduction, (2) sustainability, (3) gender & labor, (4) representation, and (5) tourist-community interaction. These topics and their sub-themes highlight scholarly interests in the use of ethnic heritage resources and poverty alleviation, local perceptions of sustainability, the politics of cultural expression, gender and labor relations, and tourist-community interaction along with local community empowerment, suggesting that ethnic tourism encompasses a wide range of interdisciplinary themes.

Keywords: Ethnic Tourism, Network Analysis, Topic Modeling, Latent Dirichlet Allocation

요약: 본 연구는 2000년부터 2025년 상반기까지 발표된 소수 민족관광 관련 SSCI급 관광학 분야 국제 학술 논문을 대상으로 연구 동향을 분석하였다. Web of Science와 Google Scholar, 그리고 SSCI 등재 관광학 저널의 공식 웹사이트를 활용하여 자료를 수집하였으며, 설정된 선별 기준에 따라 총 155편을 최종 분석 대상으로 확정하였다. 연구 방법으로는 유시넷을 활용한 네트워크 분석과 잠재 디리클레 할당 토픽 모델링을 적용하여 주요 구조적 패턴을 도출하였다. 연결·위세·매개·근접 중심성에 기반한 네트워크 분석 결과, 관광 개발과 민족 공동체와 문화가 일관되게 중심적 위치를 차지하며, 소수 민족관광 연구에서 개념적 허브

Received: November 04, 2025; 1st Review Result: December 01, 2025; 2nd Review Result: January 09, 2026
Accepted: February 10, 2026

역할을 수행하고 있음을 확인하였다. 또한 문화, 관광 영향, 관광 비즈니스, 이해관계자 집단, 소수민족, 민족마을 등이 중요한 키워드로 나타나 개발, 공동체, 문화, 정체성이 교차하는 연구 주제를 반영하였다. 토픽 모델링 분석 결과, (1) 민족 유산과 빈곤 감소, (2) 지속가능성, (3) 젠더와 노동, (4) 표현과 재현, (5) 관광객-지역사회 상호작용과 권한부여 등 다섯 가지 주요 토픽이 도출되었다.

핵심어: 민족 관광, 네트워크 분석, 토픽모델링 분석, 잠재 디리클레 할당

1. 서론

소수 민족관광(Ethnic minority tourism)은 문화관광의 한 분야로, 소수민족 지역 발전을 위해 전 세계 여러 지역에서 관광 개발이 추진되어 왔으며, 이에 따라 소수 민족관광 개발과 관련한 연구도 꾸준히 이루어져 왔다[1][2]. 선행연구는 주로 소수민족 관광지를 방문한 관광객의 인식이나 관광 개발에 대한 주민 인식을 중심으로, 소수민족 주민들의 사회경제적 취약성, 권력 불균형과 주변화 문제, 그리고 관광 개발 과정에서 드러나는 문화 재현과 전승, 주민 참여, 갈등 양상 등을 다루어 왔다[3].

그러나 이러한 연구들은 대부분 사례 연구에 머물러 전반적인 민족 관광 연구의 흐름을 조망하기에는 한계가 있다. 이러한 맥락에서 최근 문헌 고찰 연구에서는 소수 민족관광 개발의 주민 인식을 중심으로 경제적·비경제적 요인이 관광 개발 지지로 이어지는 점이 검토되었으나[4], 여전히 내용 분석 중심의 접근에 머물렀다. 따라서 소수 민족관광 연구 전반적인 동향을 검토할 수 있는 대안적 접근이 필요하다.

한편, 최근에는 네트워크 분석과 토픽모델링 등 새로운 방법론이 도입되면서 다양한 각도에서 문헌을 고찰하는 데 활용되고 있다. 지금까지 소수 민족관광 연구는 인류학·사회학적 전통에 기반한 질적 연구에 집중되어 왔으나[4], 연구 전반을 종합적으로 분석한 시도는 드물다. 예컨대 Shi, Rasoolimanesh, & Woosnam(2024)은 소수 민족 관광에 대한 주민 인식의 선행 및 후행 요인을 정리하였지만, 범위가 관광 개발 인식에 한정된 내용 분석이라는 점에서 한계를 지닌다[4]. 이에 따라 소수 민족관광 연구 전반을 포괄적으로 이해하려는 시도에는 여전히 학문적 공백이 존재한다.

이러한 학문적 공백은 소수 민족관광 연구가 관광객 인식, 주민 태도, 문화적 재현 등 다양한 주제를 다루어 왔음에도 불구하고, 연구들이 주요 개념을 중심으로 어떻게 연결·확장되어 왔는지와 주제 간 구조적 관계를 종합적으로 조망하지 못한 데서 나타난다. 즉, 기존 연구는 개별 사례와 특정 쟁점 분석에는 충실했으나 연구 성과가 분산적으로 축적되면서 연구 전체 구조 속에서 중심적 논의를 체계적으로 파악하고 이를 통합된 연구 지형으로 이해하는 데 한계를 보여 왔다.

이에 본 연구는 국제 학술지에 게재된 소수 민족관광 연구를 대상으로 네트워크 분석과 LDA 토픽모델링을 적용하여 국제 저널에서 소수 민족관광 연구가 전개되어 온 양상과 주요 논의 흐름을 검토하고자 한다. 분석 대상은 2000년부터 2025년 상반기까지 발표된 관련 논문을 대상으로 체계적 문헌고찰 절차에 따라 선별하였다. 네트워크 분석을 통해 핵심 키워드 간 관계 구조를 파악하고, 토픽모델링을 통해 잠재적 주제 구조를 도출함으로써 국제 학술지 소수 민족관광 연구 흐름을 조망한다.

2. 학술적 배경

2.1 소수 민족관광과 원주민 관광

소수 민족관광은 소수민족의 생활양식, 전통 의례, 언어, 예술, 종교, 축제 등 고유 문화를 관광자원으로 활용하여 관광객이 민족적 정체성과 생활문화를 체험하도록 하는 관광 형태를 의미한다[1]. 문화관광이 인류 문화유산과 예술 전반을 포괄한다면, 소수 민족관광은 소수민족 공동체의 문화유산과 민족적 상징의 재현에 초점을 둔다는 점에서 차별성을 지닌다[5]. 이러한 관광은 문화의 상업화와 왜곡, 정체성 갈등과 같은 문제를 동반할 수 있어 진정성 있는 문화 보존과 지속가능성이 주로 논의되어 왔다[5][6].

한편 원주민 관광은 원주민 공동체가 관광 자원의 소유·관리·운영에 직접 참여하고, 문화와 전통을 존중의 관점에서 경험하도록 하는 관광으로 정의되며[7][8], 토착성과 역사적 침탈 경험을 배경으로 공동체의 자기결정권과 통제권을 중시한다[9-11]. 두 관광 유형은 지역사회 기반의 문화적 상호작용이라는 공통점을 지니지만, 소수 민족관광은 관광객 중심성이 상대적으로 강조되는 경향이 있는 반면 원주민 관광은 공동체 주체성과 이익 통제가 보다 강하게 부각된다는 점에서 차이가 있다[10][11].

그러나 원주민 집단이 소수 집단에 포함되거나 개발 과정에서 범주가 중첩되는 경우가 적지 않아, 최근 계량서지 기반 연구에서는 소수 민족관광과 원주민 관광을 병행하여 분석하는 경향이 확인된다[4][12]. 이에 본 연구는 두 개념의 차이를 인식하되, 소수 민족관광 연구의 전반적 흐름을 파악하기 위해 ‘ethnic tourism’과 ‘indigenous tourism’을 포괄적 분석 대상으로 설정하고, 문헌을 수집하였다.

2.2 국내 소수 민족관광 연구

국내 소수 민족관광 선행연구는 중국, 베트남, 몽골 등 해외 소수민족 지역과 재외동포 대상 관광 프로그램을 중심으로 제한적으로 축적되어 왔다. 진치즈·임우호·윤지환(2023)은 중국 내몽골 소수 민족관광지를 사례로 체험관광 요소와 관광지 진정성이 관광 만족도와 행동의도에 미치는 영향을 분석하였으며[13], 반발레리아(2024)는 재외동포 모국연수 프로그램을 통해 체험경제 요인이 장소애착과 민족 정체성에 미치는 영향을 확인하였다[14]. 한편 최호림(2018)은 베트남 소수 민족관광지의 주민 역량 부족과 문화 보존-상품화 간의 긴장을 지적하며 지역기반관광과 역량개발 중심 접근의 필요성을 제시하였고[15], 정찬영·김주호(2023)는 관광두레 사례를 통해 주민 주체성, 전문인력 육성, 교육·훈련과 성과 관리 체계가 정책 실행의 핵심임을 제시함으로써 국내 소수 민족관광 연구에서 역량 강화 논의가 구체화되고 있음을 보여준다[16].

2.3 LDA (Latent Dirichlet Allocation) 토픽 모델링

LDA(Latent Dirichlet Allocation)는 문서 집합에 잠재된 주제를 확률적으로 추출하는 생성 모형으로, 하나의 문서는 여러 주제의 혼합으로 구성되며 각 주제는 특정 단어들이 함께 출현하는 확률적 패턴으로 표현된다는 가정에 기반한다[17]. 이 모형을 통해 단어 분포와 주제 분포를 학습함으로써 자료에 내재된 주제 구조와 통계적 특성을 파악할 수 있다. 토픽 모델링에서 최적 주제 수는 혼란도(perplexity), 일관성(coherence), 로그 우도, 주제 간 상관관계, 하이퍼파라미터(α) 최적화 등 다양한 지표를 통해 결정되어 왔으며[12], 본

연구는 관광 분야 선행연구를 토대로 혼란도와 일관성 지수를 기준으로 최적 주제 수를 산정하였다[18]. 혼란도는 모델의 예측 성능을 나타내는 지표로 값이 낮을수록 적합도가 높음을 의미하며, 일관성은 각 주제 내 상위 단어들의 의미적 응집성을 평가한다[12].

2.4 네트워크 분석

소셜 네트워크 분석(Social Network Analysis)은 의사소통 집단이나 관계망 속 개체들의 상호작용을 정량적으로 파악하고, 이들의 연결 구조를 시각적으로 표현하는 기법이다[19]. 이수상(2014)은 이를 특정 노드(예: 개인, 지식, 장소, 조직 등) 간 관계를 파악하는 이론적 틀로 설명하며, 노드와 링크를 통해 네트워크 구조를 모델링 결과, 중요한 위치에 있는 개체를 도출할 수 있다고 보았다[19]. 소셜 네트워크 분석을 확장한 방법이 의미·키워드·텍스트 네트워크 분석 등으로도 불린다[20]. 텍스트 네트워크 분석을 “단어를 노드로 정의하고, 단어의 동시 출현 관계를 링크로 코딩하여 텍스트 내 의미 구조를 추출하는 기법”으로 정의하였다[19]. 이러한 접근은 단어 빈도 분석을 넘어 문서 전체의 의미 구조와 네트워크상 위치와 영향력 분석을 목적으로 한다.

3. 연구 방법

3.1 자료 수집과 연구 문제

본 연구는 2000년부터 2025년 상반기까지 해외 주요 관광 저널에 게재된 소수 민족관광 연구의 전반적 동향을 분석하는 것을 목적으로 한다. 이를 토대로 다음과 같은 연구 문제를 설정하였다. 첫째, 소수 민족관광 연구의 시기별·저널별 게재 분포를 검토한다. 둘째, 네트워크 중심성 분석을 통해 핵심 키워드와 그 연계 구조를 규명한다. 셋째, 토픽모델링 분석을 통해 소수 민족관광 연구의 주요 주제와 특성을 탐색한다.

본 연구는 자료 식별-제외 기준의 동시 적용-개별 검토-최종 포함의 절차에 따라 문헌고찰을 수행하였다[21]. 자료 식별은 2025년 6월 초 Web of Science(WoS) 검색에서 시작하였으며, Topic Search = (“ethnic tourism” OR “indigenous tourism” OR “ethnic minority tourism” OR (“ethnic minority” AND “tourism”))의 Boolean 검색식을 적용하여 제목·초록·키워드 영역에서 문헌을 수집하였다. 주요 검색어는 “ethnic tourism”, “indigenous tourism”, “ethnic minority tourism”이다.

WoS 1차 검색 이후 관련 논문의 제목을 확인하고, 각 논문 제목을 기준으로 Google Scholar(GS)를 병행 검색하였다[22]. 검색 직후 수집된 문헌은 즉시 엑셀 파일에 정리하였다. 또한 검색 과정에서 확인된 관광학 분야 SSCI 등재 주요 저널의 공식 웹사이트에서도 동일 검색어를 활용하여 추가 식별을 수행함으로써 잠재적 누락 가능성을 최소화하고자 하였다. 이 과정에서 확인된 문헌 역시 엑셀 파일에 체계화하였다. 이와 같은 반복 검색 및 정리 절차를 거쳐 엑셀 데이터베이스를 축적한 이후, GS에 동일한 검색어 조합을 적용하여 각 검색어별 상위 20페이지까지 검토하였다. 이를 통해 자료 수집 과정에서 발생할 수 있는 누락 가능성을 최소화하고자 하였다.

한편, 문헌 선별 기준은 선형적 단계 구분이 아니라 복수 기준을 병행 적용하는 방식으로 이루어졌다. WoS 검색 단계에서 문헌 유형(회의논문 제외), 언어(비영어권 제외), 학문 분야(비관광 분야 제외) 필터를 1차적으로 적용한 후, 각 문헌의 제목과 저널명을 연구자가 개별적으로 확인하여 SSCI 등재 여부를 재검토하였다. GS를 통해 수집된 문헌

역시 연구자가 문헌 유형, 언어, 학문 분야, 저널 등재 여부 및 SSCI 여부를 개별적으로 확인하는 절차를 병행 적용하였다. 이러한 제외 기준은 특정 단계에서 일괄적으로 적용된 것이 아니라 전 과정에서 동시에 고려되었으며, 최종적으로 제목·초록·주제 적합성에 대한 연구자의 개별 평가를 거쳐 분석 대상을 확정하였다.

또한 수집된 모든 문헌은 검색 직후 즉시 엑셀 기반 데이터베이스에 수작업으로 체계화하였다. 엑셀 데이터베이스에는 제목, 초록, 저자, DOI, 연도, 저널명, APA 형식 참고문헌, 영문 초록에 대한 한국어 번역, 한국어 제목, 분류 키워드, SSCI 여부 재확인, 제목·초록 통합 검토란 등의 항목을 포함하여 서지 정보와 주제 적합성을 통합 관리하였다. 입력된 자료는 알파벳 순으로 정렬하였으며, 이후 제목·저자·DOI를 기준으로 개별 검토를 수행하여 중복 문헌을 제거하였다.

문헌 검색은 2025년 6월 초부터 6월 26일까지 진행되었으며, 제목과 초록을 중심으로 *ethnic tourism* 관련 연구 여부를 검토하여 최종 선별을 완료하였다. 그 결과, 2025년 6월 26일 기준 총 155편의 논문을 분석 대상으로 확정하였다. 이는 관광개발 맥락에서 민족관광 연구를 대상으로 범위를 설정한 Shi, Rasoolimanesh, & Woosnam(2024)의 메타연구에서 선정된 64편과 비교할 때 상대적으로 많은 규모에 해당한다.

3.2 분석 방법

네트워크 분석을 위해 키워드 공출현 행렬을 구축하고 UCINET 6.812를 사용하여 중심성을 산출하였으며, 연결 강도를 반영한 비방향성(undirected) 네트워크로 분석을 수행하였다. 분석의 신뢰성을 확보하기 위해 동의어 및 유사 키워드 통합, 불용어 제거 등의 전처리를 실시하였다. 본 연구에서는 최소 출현 빈도를 사전에 임의로 설정하지 않았다. 대신 전체 키워드를 출현 빈도 순으로 정렬한 후, 네트워크 구조의 안정성과 해석 가능성을 고려하여 상위 빈도 키워드를 분석 대상으로 선정하였다. 그 결과 최종 포함된 키워드 집합의 최소 출현 빈도는 2회로 나타났으며, 이는 데이터 분포에 기반한 선별의 결과이다. 공출현은 문헌 단위에서 초록을 기준으로 계산하였다.

토픽모델링 분석은 Python 3.8 환경에서 수행되었으며, 주요 라이브러리를 활용하여, 초록 텍스트에 대해 특수기호 제거와 단어 정규화를 거친 뒤, 형태소 분석기를 사용해 명사를 추출하고 불용어 및 1글자 단어를 제거하였다. 이후 코퍼스와 사전을 구축하여 단어 전처리 정제 및 표준화 이후, 문서를 Bag-of-Words 형태로 변환하였다. 토픽 수는 3~9 범위에서 혼란도(Perplexity)와 일관성(Coherence Score)를 기준으로 탐색하여 최적 값을 선정하였으며, Gensim의 LDA 모델을 반복 학습해 토픽별 핵심 키워드를 도출하고 결과를 시각화하였다. 토픽 개수의 설정은 분석의 정확성과 해석 가능성을 동시에 고려해야 하는 중요한 과정이다. 토픽 수가 지나치게 적을 경우 서로 다른 주제가 하나의 토픽에 포함되어 주제 간 구분이 모호해질 수 있으며, 반대로 지나치게 많을 경우 유사한 내용이 여러 토픽으로 분할되는 과군집화(over-clustering) 현상이 발생할 수 있다. 이러한 문제를 최소화하기 위해 본 연구에서는 Coherence(응집도)와 Perplexity(혼란도) 지표를 함께 활용하였다.

Coherence 지표는 Gensim에서 제공하는 *c_v* 방식을 사용하여 계산하였으며, 이는 토픽 내 단어 간 공출현에 기반한 의미적 일관성을 측정하는 대표적인 지표이다. 값이 높을수록 명확하고 해석 가능한 주제를 반영하는 것으로 이해된다. Perplexity는 모델의 예측 성능을 평가하는 지표로, 값이 낮을수록 문서 분포를 보다 잘 설명하는 것으로

해석된다. 그러나 Perplexity 값이 낮다고 해서 반드시 해석 가능성이 높다고 볼 수는 없다. 이에 따라 본 연구에서는 Perplexity와 Coherence(c_v)를 함께 고려하여 최적 토픽 수를 결정하였다.

LDA 분석은 $\alpha=0.1$, $\beta(\eta)=0.01$, $\text{passes}=50$, $\text{iterations}=1000$ 의 설정으로 수행하였으며, 최종 보고 모델에는 $\text{random seed}=123$ 을 적용하였다. 한편, LDA는 확률적 특성을 지니는 비결정론적 알고리즘이므로 초기값에 따라 결과가 일부 달라질 수 있다. 이에 모델의 안정성을 검증하기 위해 서로 다른 random seed 값을 적용하여 다중 반복 학습 실행을 수행하였다. 동일한 전처리 조건과 토픽 수 설정 하에 반복 실행한 결과를 상호 비교하여 토픽 분포와 상위 키워드 구성이 일관되게 유지되는지를 확인함으로써, 특정 초기화에 따른 편향 가능성을 최소화하고 토픽 구조의 안정성을 검증하였다.

3.3 텍스트 전처리 및 용어 정규화

본 연구는 1-gram 및 2-gram 단어를 대상으로 전처리를 수행하였다. 전처리는 표현 순서의 통일, 일부 표현의 정리, 관광 맥락의 명확화, 그리고 형태 표준화의 원칙에 따라 이루어졌다. 먼저, 표현 순서의 통일이다. *development tourism*과 *tourism development*, *impact tourism*과 *tourism impact*처럼 단어 배열만 상이한 경우는 각각 *tourism_development*, *tourism_impact*로 통일하여 동일 개념이 분산되지 않도록 하였다. 다음으로, 일부 표현의 정리이다. *poverty alleviation*과 *poverty_reduction* 등은 *poverty_reduction*으로 표기 통일하였으며, *woman entrepreneur*는 *ethnic_woman_tourism_worker*로 정리하였다.

또한 관광 맥락의 명확화를 위하여 관광 담론에서 반복적으로 사용된 단어 결합은 분석 단위를 일관되게 유지하도록 정규화하였다. 예컨대 *tourist interaction*과 *community interaction*은 *tourist_community_interaction*으로, *community member*와 *member group*은 *community_member*로 표기하였다. *government policy*는 *government_tourism_policy*로 정리하였다. 아울러 *community tourism*은 문헌에서 *community-based tourism*이라는 개념으로 사용되는 점을 반영하여 *community_based_tourism*으로 정리하였다.

마지막으로, 형태 표준화이다. 지리적 표기의 일관성을 위해 *yunnan province*, *yunnan china*, *destination yunnan* 등은 *yunnan_province*로 통일하였고, *labour*는 *labor*로 표준화하였다.

또한 토픽 형성에 직접적으로 기여하지 않는 일반 학술 용어는 제거하였다. 연구 수행을 설명하는 형식적 용어(예: *study*, *research*, *paper*, *article*), 결과 보고와 관련된 표현(예: *result*, *finding*, *discussion*, *literature*), 방법론적 용어(예: *method*, *analysis*, *questionnaire*, *sample*), 그리고 의미 범위가 지나치게 포괄적인 추상적 단어(예: *aspect*, *dimension*, *factor*, *concept* 등)가 포함되었다. 이러한 전처리는 분석 단위의 일관성을 유지하고 토픽 해석의 명확성을 확보하기 위한 정규화 과정에 해당한다.

4. 분석 결과

4.1 시기별·저널별 현황

2000년부터 2013년까지는 해외 소수 민족관광 연구가 연간 1~7편 수준으로 산발적으로 발표되며 전반적으로 소규모에 머물렀다. 2014년 이후에는 연구가 점차 확산되어 2016년에는 처음으로 두 자릿수(11편)에 도달하였고, 2019년과 2020년에도 각각 12편과 11편으로 높은 수준을 보였다. 2021년에는 3편으로 일시적으로 감소했으나, 2022년 이후

다시 증가하여 2024년에는 13편으로 가장 활발하게 발표되었고, 2025년에도 10편을 기록하며 일정한 연구 흐름을 유지하였다.

한편, 소수 민족관광 연구는 Journal of Tourism and Cultural Change가 전체의 17.95%(21편)를 차지하며 가장 높은 비중을 기록하였다. 이어서 Journal of Sustainable Tourism(13편, 11.11%), Asia Pacific Journal of Tourism Research(11편, 9.40%), Annals of Tourism Research(10편, 8.55%), Tourism Management(10편, 8.55%) 등이 뒤를 잇고 있어, 소수 민족관광 연구가 주로 관광학 및 지속가능성 관련 대표 학술지에 발표되었음을 알 수 있다. 또한 Sustainability(8편, 6.84%), Current Issues in Tourism(7편, 5.98%), Journal of Hospitality and Tourism Management(6편, 5.13%), Tourism Geographies(5편, 4.27%)에서도 꾸준히 논문이 게재되었다. 반면, Journal of Heritage Tourism, Tourism Management Perspectives, International Journal of Tourism Research 등은 각각 4편(3.42%)을 수록하며 중위권을 형성하였다. 이러한 결과는 일부 저널에 집중적으로 축적되면서도, 다양한 주변 학술지를 통해 점차 확산되고 있음을 보여준다.

4.2 네트워크 분석 결과

본 연구에서는 소수 민족관광 관련 주요 개념 간 관계를 파악하기 위해 네트워크를 구성하고 구조 지표와 중심성 지표를 산출하였다. 본 연구에서 구성된 네트워크는 50개의 노드와 1,822개의 연결로 이루어졌으며, 밀도는 0.744로 나타났다. 네트워크는 단일 컴포넌트 구조를 보였고(Components=1), 연결성은 1, 분절성은 0으로 확인되었다. 평균 경로 길이는 1.256, 직경은 2로 산출되었다. 다음 각 개념이 네트워크 내에서 차지하는 상대적 위치를 파악하기 위해 중심성 분석을 실시하였다. 연결 중심성은 빈도 기반의 양적 중요성, 위세 중심성은 질적 영향력, 매개 중심성은 주제 간 연결성을 반영한다. 분석 결과, tourism development, community, culture, tourism impact, tourism business, stakeholder 관련 단어들이 중심성 지표 모두에서 상위 10개에 위치하여, 소수 민족관광 연구 동향 네트워크의 구조적·개념적 중심축을 이루는 핵심어임이 확인되었다([표 1] 참조).

[표 1] 네트워크 중심성 지표별 상위 10개 키워드

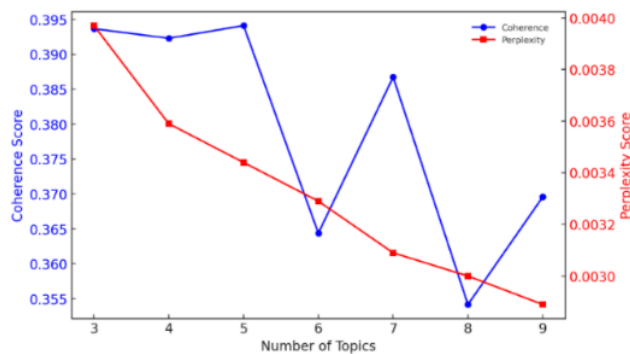
[Table 1] Top 10 Results by Network Centrality

Rank	Degree (연결)	Eigenvector (위세)	Betweenness (매개)	Closeness (근접)
1	tourism development (419)	tourism development (0.418)	tourism development (15.444)	ethnic neighborhood tourism (81)
2	community (372)	community (0.379)	community (15.444)	medium (76)
3	culture (216)	culture (0.233)	culture (12.664)	tourism support (72)
4	tourism impact (213)	tourism business (0.231)	stakeholder group (12.246)	poverty reduction (72)
5	tourism business (211)	tourism impact (0.223)	tourism area (11.865)	australia (71)
6	stakeholder group (169)	stakeholder group (0.191)	tourist (11.655)	community participation (71)
7	tourist (168)	community based tourism (0.184)	tourism impact (11.391)	livelihood (70)
8	community based tourism (163)	tourist (0.177)	tourism business (11.216)	tourist community interaction (70)

아울러 tourism impact, tourism business, stakeholder group, community-based tourism 등의 개념이 주요 위치를 차지함으로써, 소수 민족관광 연구가 관광 산업, 이해관계자 구조, 공동체 기반 접근이 결합된 담론으로 전개되고 있음을 보여준다. 한편, 근접 중심성 분석에서는 community participation, livelihood, poverty reduction이 상대적으로 두드러지게 나타났다.

4.3 토픽모델링 분석 결과

최적 토픽 수를 결정하기 위해 일관성(Coherence)과 혼란도(Perplexity) 점수를 함께 분석하였다. 토픽 수를 단계적으로 변화시키며 Gensim의 LdaMulticore 클래스를 이용해 모델을 학습한 결과([그림 1] 참조), 일관성 점수는 토픽 수 5에서 0.39413으로 가장 높게 나타났다. 반면, 혼란도 점수는 토픽 수 증가에 따라 감소하되 7 이후부터는 감소 폭이 완만해지는 경향을 보였다. 주제 간 구분력을 확보하면서 과도한 토픽 분할로 인한 복잡성을 피하기 위해, 일관성 점수가 최고점을 기록한 5개 토픽을 최적 토픽 수로 선정하였으며 그 구체적인 결과는 다음과 같다([표 2] 참조).



[그림 1] 일관성 & 혼란도

[Fig. 1] Coherence & Perplexity

[표 2] 토픽모델링 분석 결과

[Table 2] LDA Topic Modeling Results

Topic1 (n=29, 18.71%)	Topic2 (n=23, 14.84%)	Topic3 (n=29, 18.71%)	Topic4 (n=34, 21.94%)	Topic5 (n=40, 25.81%)
heritage (0.072)	sustainability (0.041)	woman gender (0.057)	representation (0.032)	interaction (0.047)
poverty reduction (0.039)	tour guide (0.034)	labor division (0.039)	language (0.030)	tourist community interaction (0.031)
empowerment (0.025)	indicator (0.033)	income (0.033)	canada (0.026)	policy (0.028)
ethnic woman tourism worker (0.025)	tourist experience (0.033)	ethnic group (0.027)	attitude (0.026)	tourism support (0.028)
ethnic group (0.024)	interest (0.032)	ethnic woman tourism worker (0.027)	state (0.023)	capacity (0.028)

indigenous (0.024)	resident perception (0.022)	governance (0.026)	community member (0.023)	empowerment (0.027)
agency (0.021)	role (0.021)	ethnic neighborhood tourism (0.024)	discourse (0.023)	Australia (0.024)
market (0.021)	conflict (0.019)	process (0.023)	evolution (0.022)	community empowerment (0.021)
Ethnic minority area (0.021)	yunnan province (0.019)	society (0.023)	ethnic neighborhood tourism (0.020)	community member (0.021)
initiative (0.021)	satisfaction (0.019)	household (0.021)	elite (0.020)	government tourism policy (0.019)

주: n은 각 토픽에 분류된 문서 수를 의미하며, Percentage(%)는 전체 문서(N=155) 대비 비율임.

Topic 1은 heritage(0.072)를 중심으로 poverty reduction, empowerment, ethnic woman tourism worker가 함께 출현하여 유산-빈곤감소-역량강화-소수민족 여성 논의를 보였다. Topic 2는 sustainability를 축으로 tour guide, indicator, tourist experience, interest가 나타나 지속가능성-지표-가이드-관광객 경험 담론의 결집을 보여준다. Topic 3은 woman gender, labor division, income, ethnic group, governance의 동시 출현으로 젠더-노동 분업-소득 논의가 드러났다. Topic 4는 representation, language, Canada, attitude, state가 함께 나타나 재현-언어-국가 맥락의 결합을 시사하였다. Topic 5는 interaction, tourist community interaction이 policy, tourism support, capacity와 동반되어 상호작용-정책-지원-역량 구축 프레임이 관찰되었다.

5. 결론 및 시사점

본 연구는 소수 민족관광 연구동향을 분석하기 위해 네트워크 분석과 LDA 토픽모델링을 적용하였다. 네트워크 분석 결과, 연결·위세·매개 중심성 지표에서 공통적으로 상위에 위치한 키워드가 도출되었으며, 이는 소수 민족관광 연구 담론이 관광 개발과 영향, 소수민족 커뮤니티와 이해관계자, 문화, 관광객을 핵심 축으로 형성되어 있음을 보여준다. LDA 분석에서는 ‘문화유산과 빈곤 감소’, ‘지속가능성’, ‘젠더와 노동’, ‘문화적 재현’, ‘상호작용과 역량 강화’의 다섯 개 주요 주제가 도출되었다. 이는 소수민족 문화유산 기반 관광을 중심으로, 미시적으로는 관광객-공동체 상호작용과 주민 역량 강화 전략을, 거시적으로는 지속가능한 관광 개발이 공동체에 미치는 영향, 문화 재현, 성별과 노동, 빈곤 감소 등 복합적 쟁점을 다층적으로 논의하고 있음을 시사한다.

중심성 상위 키워드를 기준으로 LDA 토픽과 교차 분석한 결과, tourism development는 직접적으로 토픽 모델링 분석의 상위어에 포함되지는 않았으나 poverty reduction(Topic 1), sustainability(Topic 2), income(Topic 3), policy(Topic 5) 등으로 분화되어 각 토픽에 분산적으로 나타났다. 이는 개발 담론이 단일 개념으로 제시되기보다는 세부 성과 및 정책 차원으로 재구성되어 나타남을 의미한다. 또한 community는 empowerment(Topic 5), tourist community interaction(Topic 5), community member(Topic 4) 등의 형태로 분포하며 공동체 중심 접근이 핵심 담론으로 작동함을 보여준다. 한편 culture는 heritage(Topic 1)와 representation, language, discourse(Topic 4)를 통해 문화유산적 접근과 담론적 접근으로 구분되어 나타났다. 아울러 stakeholder group은 governance(Topic 3) 및 policy(Topic 5)와 연결되어 거버넌스·정책 중심 담론과 결합되는 구조를 보였다. 이러한 결과는 네트워크 중심 개념이 토픽 구조 내에서 세부 차원으로 재구성되어 나타남을 보여준다.

이는 네트워크 분석과 LDA 토픽모델링이 상이한 분석 차원에서 자료를 조명하는 보완적 접근임을 시사한다. 네트워크 분석은 키워드 간 공출현 관계를 기반으로 개념 간

연결 구조를 도출함으로써 담론의 구조적 중심성 역할을 규명하고, 개념이 관계망 속에서 차지하는 위치를 드러낸다. 반면 LDA는 단어의 확률적 분포를 토대로 잠재된 주제를 추출하여 의미 단위로서의 담론 구성을 식별한다. 즉, 네트워크 분석이 담론의 관계적 구조를 조명한다면, LDA는 담론의 의미적 구조를 제시한다고 할 수 있다. 구조적 허브로 도출된 개념이 LDA 토픽에서도 다양한 형태로 분산되어 나타나는 현상은, 해당 개념이 연구 담론 전반에 폭넓게 연관되어 있음을 시사한다.

본 연구의 이론적 기여는 소수 민족관광 담론을 개별 주제의 빈도 분류 수준을 넘어, 개념 간 결합 양상과 잠재적 의미 구조로 재해석하였다는 점에 있다. Shi, Rasoolimanesh, & Woosnam(2024)은 민족관광에 대한 64편의 연구를 분석하여 Conflicts/confrontations(12편), Ethnic values(10편), Community participation(8편), Tourism impacts(6편), Quality of life/well-being(5편), Host-guest interaction(5편), Gender labour division/empowerment(5편), Sustainable livelihood/development(4편), Place attachment(3편), Authenticity(2편), Cultural/ethnic identity(2편), Emotional solidarity(1편), Community support(1편) 등 13개 주제를 도출하고, 이를 경제·사회문화·환경 및 사회정치 차원으로 통합함으로써, 소수 민족관광 연구가 갈등과 가치·정체성 중심으로 전개되어 왔음을 정리하였다. 그러나 이러한 접근은 개념 간 결합 구조까지 설명하는 데에는 제한이 있다.

반면 본 연구는 LDA 기반 토픽모델링을 활용하여 5개 토픽의 문서 분포와 키워드 확률값을 토대로 담론의 잠재 구조를 분석하였다. Shi et al.(2024)에서 Host-guest interaction과 Gender labour division/empowerment가 각각 독립 범주로 제시되었다면, 본 연구에서는 상호작용과 관련하여 정책·역량 강화와 결합된 Topic5(interaction-policy-capacity)로, 젠더는 노동 분화·소득 구조와 결합된 Topic3(gender-labor division-income)로 나타났으며, Topic1(heritage-poverty reduction)은 Shi et al.(2024)의 Ethnic values/identity 및 Sustainable livelihood/development 논의와 개념적 접점을 형성하였다. 이는 관계, 젠더, 유산 논의가 정책·역량·경제 구조와 교차하는 담론 층위에서 구성되고 있음을 보여준다.

또한 네트워크 중심성 분석을 결합한 결과, tourism development는 중심 위치에, stakeholder group은 매개 위치에, ethnic neighborhood tourism은 확산 가능성이 높은 위치에 나타났다. 이러한 분석은 주제의 분포 정리에 머무르지 않고, 개념 간 결합 패턴과 구조적 위치를 통해 담론의 형성 방식을 설명한다는 점에서 분석 차원을 확장한다.

나아가 국제 소수 민족관광 연구를 대상으로 도출된 주제 구조는 국내 소수 민족관광 연구에 참고 가능한 비교 관점을 제공한다. 국제 연구에서 반복적으로 논의된 문화유산 활용, 지속가능성, 지역사회-관광객 상호작용, 공동체 역량 강화는 인구 감소와 다문화 이주가 확산되는 한국 사회 맥락에서 향후 지역 문화관광 연구의 주제 설정과 개발 논의 방향을 모색하는 데 활용될 수 있다.

본 연구의 결과는 소수 민족관광 연구의 담론 구조를 제시하는 탐색적 분석으로서 실무적 함의를 지닌다. 도출된 토픽들은 소수 민족관광 개발 정책에서 문화유산 보존과 지역경제 발전의 조화, 지속가능성 확보, 젠더와 노동의 사회적 영향, 문화적 재현, 주민-관광객 상호작용, 주민 역량 강화를 핵심 논점으로 제시한다. 분석 결과, 수익 중심의 개발을 넘어 지역사회 가치와 문화유산 보호를 병행할 필요가 있으며(토픽 1·2), 주민 인식에 기반한 정책 설계와 이해관계자 협력은 갈등 관리와 관광객 만족도 제고를 통한 지속가능성 확보의 핵심 과제로 도출되었다(토픽 5). 나아가 관광이 공동체 정체성과 사회적 변화를 수반할 수 있음을 고려할 때, 성별을 포괄한 역량 강화 정책과 소수민족 언어를 정체성과 장소성의 핵심 자원으로 인식하는 주민 참여 중심의 접근이

요구된다(토픽 3·4)[15][16].

본 연구는 몇 가지 한계를 지닌다. 첫째, 특정 키워드와 영문 논문에 한정된 문헌 수집으로 인해 유산관광·생태관광 등 인접 연구와 비영어권 연구가 충분히 반영되지 못해 지역적 대표성에 제약이 있다. 또한 국제 연구와 국내 연구를 병렬적으로 분석함으로써 연구 결과의 맥락적 적용 가능성을 보다 구체화할 필요가 있다. 둘째, 토픽모델링은 전처리, 매개변수 설정, 주제 해석 과정에서 연구자의 판단이 개입될 수 있어 결과의 객관성이 일부 제한된다. 셋째, 본 연구는 주요 주제 도출에 초점을 두어 지역별 차이, 이론·방법론의 비교, 담론의 심층 분석까지 포괄하지 못하였다. 또한 소수 민족관광 연구 담론의 변화를 시기별·주제별 단계 모델로 체계화하여 제시하지 못한 점은 본 연구의 한계로 남는다. 주요 주제의 부상과 약화, 담론 전환 구간을 시계열적으로 구분하는 분석까지는 수행하지 못하였다는 점에서, 연구 지형의 발전 단계를 이론적으로 추상화하는 데 일정한 제약이 있었다.

향후 연구에서는 동적·구조적 토픽모델링을 적용하여 담론 변화를 시기별·주제별·집단적 차원에서 체계화할 필요가 있다. 예컨대 연도별 자료를 구분한 동적 토픽모델링(Dynamic Topic Modeling)을 활용하면 특정 시점에서 강화되거나 약화되는 주제를 추적할 수 있으며, 지역 등 공변량을 반영한 구조적 토픽모델링(Structural Topic Modeling)은 시기별 주제 비중의 통계적 차이와 담론 전환 구간을 검증하는 데 유용하다. 더 나아가 시기별 키워드 네트워크를 구축하여 중심성 지표의 변화를 비교한다면, 구조적 허브 개념의 이동 여부를 통해 담론의 구조적 변동을 분석할 수 있을 것이다. 이러한 확장적 접근은 소수 민족관광 연구의 발전 단계를 보다 정교하게 구분하고, 연구 지형의 변화 과정을 체계적으로 설명하는 데 기여할 것으로 기대된다.

또한 본 연구는 자료 식별-제외 기준의 동시 적용-개별 검토-최종 포함의 절차에 따라 문헌을 확정하였으나, 각 단계별 논문 수를 PRISMA 흐름도 형식으로 세분화하여 제시하지는 못하였다[21]. 이에 따라 자료 검색, 중복 제거, 제목·초록 검토, 전문 검토, 최종 포함 단계별 수치를 체계적으로 시각화하지 못한 한계가 존재한다.

향후 연구에서는 PRISMA 가이드라인에 부합하도록 초기 검색 건수, 중복 제거 후 건수, 제외 사유별 배제 건수, 최종 포함 건수 등을 단계별로 구체화하고, 배제 사유 또한 범주화하여 명확히 보고할 필요가 있다. 이를 통해 문헌 검색 및 선별 과정의 투명성과 재현 가능성을 보다 강화할 수 있을 것이다.

또한 다중 데이터베이스를 활용하는 경우에는 데이터베이스별 초기 검색 결과와 중복 제거 과정, 최종 통합 수치를 구분하여 제시함으로써 체계적 문헌고찰의 보고 기준을 충실히 반영할 필요가 있다. 아울러 향후 연구에서는 상업적 데이터베이스와 함께 OpenAlex와 같은 오픈소스 기반 학술 인프라를 병행 활용하고, 검색 범위(연도, 언어, DOI 보유 여부, 학문 분야, 문헌 유형 등)와 필터 조건을 명확히 공개함으로써 외부 검증 가능성을 높일 필요가 있다. 또한 데이터베이스 간 검색 결과의 차이를 비교·보고함으로써 문헌 수집 범위의 투명성과 학술적 신뢰성을 더욱 제고할 수 있을 것이다.

References

- [1] J. Yang, G. Wall, Ethnic tourism: A framework and an application, *Tourism Management*, (2009), Vol.30, No.4, pp.559-570.

DOI: 10.1016/j.tourman.2008.09.008

- [2] K. Ishii, The impact of ethnic tourism on hill tribes in Thailand, *Annals of Tourism Research*, (2012), Vol.39, No.1, pp.290-310.
DOI: 10.1016/j.annals.2011.05.004
- [3] Y. Wang, H. Shen, S. Ye, L. Zhou, Being rational and emotional: An integrated model of residents' support of ethnic tourism development, *Journal of Hospitality and Tourism Management*, (2020), Vol.44, No.2, pp.112-121.
DOI: 10.1016/j.jhtm.2020.05.008
- [4] S. J. Shi, S. M. Rasoolimanesh, K. M. Woosnam, A systematic review of residents' perceptions towards tourism development in ethnic tourism contexts, *Journal of Tourism and Cultural Change*, (2024), Vol.22, No.4, pp.382-406.
DOI: 10.1080/14766825.2024.2401838
- [5] L. Yang, Ethnic tourism and cultural representation, *Annals of Tourism Research*, (2011), Vol.38, No.2, pp.561-585.
DOI: 10.1016/j.annals.2010.10.004
- [6] L. Yang, G. Wall, Authenticity in ethnic tourism: Domestic tourists' perspectives, *Current Issues in Tourism*, (2009), Vol.12, No.3, pp.235-254.
DOI: 10.1080/13683500802406880
- [7] R. Aulia, R. Kusumastuti, Indigenous tourism concept of building friendly tourism for indigenous Baduy people: A literature study, *Budapest International Research and Critics Institute-Journal*, (2022), Vol.5, No.2, pp.11026-11032.
DOI: 10.33258/birci.v5i2.4920
- [8] Tourism Australia, Indigenous tourism: A way to be immersed in the culture and nature of a country, *Tourism Australia*, (2008)
Available from: <https://www.tourismdashboard.org/indigenoustourism-a-way-to-be-immersed-in-the-culture-and-nature-of-a-country>
- [9] A. Sharma, Indigenous and tribal tourism, Inclusive community development through tourism and hospitality practices, *IGI Global*, (2023)
DOI: 10.4018/978-1-6684-6796-1.ch005
- [10] J. Zhou, J. R. Edelman, Ethnic minority tourism and indigenous tourism: The critical distinction, *Tourism Culture & Communication*, (2023), Vol.23, No.1, pp.1-12.
DOI: 10.3727/109830422X16510695152064
- [11] A. Carr, L. Ruhanen, M. Whitford, Indigenous peoples and tourism: the challenges and opportunities for sustainable tourism, *Journal of Sustainable Tourism*, (2016), Vol.24, No.8-9, pp.1067-1079.
DOI: 10.1080/09669582.2016.1206112
- [12] Y. Cai, Y. Xiao, J. Zhou, Z. Gong, Decoding ethnic tourism: A comprehensive analysis of global trends, key themes, and knowledge frameworks, *Humanities and Social Sciences Communications*, (2025), Vol.12, No.1, pp.1-18.
DOI: 10.1057/s41599-025-05769-w
- [13] C. Z. Jin, W. H. Lim, J. H. Yoon, The impact of experience tourism elements and authenticity of attractions in China's ethnic minority regions on tourist satisfaction and behavioral intention, *International Journal of Tourism and Hospitality Research*, (2023), Vol.37, No.8, pp.63-76.
DOI: 10.21298/IJTHR.2023.12.37.8.63
- [14] V. Pan, How does tourism experience transform the ethnic identity of compatriots? Focusing on Experience Economy Theory (4Es), Hanyang University Graduate School, Master's Thesis, (2024)
Available from: <https://repository.hanyang.ac.kr/handle/20.500.11754/189468>
- [15] H. R. Choi, Capacity Building for Sustainable Ethnic Tourism Development in Vietnam, *The Journal of Asian Studies*, (2018), Vol.21, No.3, pp.23-49.
DOI: 10.21740/jas.2018.08.21.3.23
- [16] C. Y. Jung, J. H. Kim, A Study on the Development of Professional Personnel for Promoting Tourism for Minorities: Focusing on the Case of "Tourism Dure", *Korean Journal of Siberian Studies*, (2023), Vol.27, No.2, pp.67-90.
DOI: 10.22892/ksc.2023.27.2.03

- [17] A. P. Kirilenko, S. Stepchenkova, Facilitating topic modeling in tourism research: Comprehensive comparison of new AI technologies, *Tourism Management*, (2025), Vol.106, 105007.
DOI: 10.1016/j.tourman.2024.105007
- [18] H. Gomin, Analysis of domestic cruise tourism research trends using the LDA topic modeling technique, *Asia-Pacific Journal of Convergent Research Interchange*, (2024), Vol.10, No.9, pp.167-178.
DOI: 10.47116/apjcri.2024.09.16
- [19] S. Lee, Content analysis of academic papers using language network analysis methods, *Journal of the Korean Society for Information Management*, (2014), Vol.31, No.4, pp.49-68.
DOI: 10.3743/KOSIM.2014.31.4.049
- [20] S. Yoo, K. Lim, News agenda analysis using text mining and semantic network analysis: Focusing on emotions related to COVID-19, *Journal of Intelligence and Information Systems*, (2021), Vol.27, No.1, pp.47-64.
DOI: 10.13088/jiis.2021.27.1.047
- [21] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, D. Moher, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration, *BMJ*, (2009), Vol.339, b2700.
DOI: 10.1136/bmj.b2700
- [22] A. Martín-Martín, E. Orduna-Malea, A. W. Harzing, E. Delgado López-Cózar, Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories, *Journal of Informetrics*, (2018), Vol.12, No.4, pp.1160-1177.
DOI: 10.1016/j.joi.2018.09.002