

# 북극 소수민족에 대한 러시아 언론의 보도내용 및 성향 분석: ‘타스 통신사’의 뉴스 기사 텍스트를 중심으로\*

계용택\*\*

| 목 차 |

I. 들어가는 말	IV. 연구방법
II. 선행연구 분석	V. 연구내용 및 결과
III. 자료수집	VI. 결론

| 논문요약 |

본 논문은 러시아에서 발행되는 뉴스를 분석하여 러시아 국민 및 정부가 북극에 거주하는 소수민족들에 대한 생각을 고찰하고자 하였다. 뉴스기사 텍스트 마이닝 기법에서 R 소프트웨어의 TM 패키지를 이용하였으며, 내용 분석을 위하여 단어 빈도수(워드클라우드 제시) 분석, 단어 간의 네트워크 분석, 상관분석(Correlation Analysis), 엔그램(n-gram, 바이트그램 및 트라이그램) 분석 방법 등으로 ‘북극 및 민족’으로 검색된 ‘타스 통신사’ 뉴스 기사 텍스트 본문을 분석하여 기사 내용의 본질적인 의도를 추론하였다.

분석결과 북극토착소수민족 단어들의 상관도는 북극개발 프로젝트에 비하면 매우 낮은 수준이라 볼 수 있다. 이러한 연구결과로 볼 때 전체 기사의 핵심 키워드는 ‘러시아 북극지역 사회경제적 개발’ 및 ‘북극지역 소수민족 언어 권리’이며, 전체 텍스트 문맥상 북극의 소수민족 문제는 러시아의 북극개발 과정에 있어 아주 작은 부분에 지나지 않다는 점을 보여주고 있다. 북극토착소수민족들에 대한 러시아 국민 및 정부의 태도를 감안하면, 이러한 정책시행으로 북극에 거

\* 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5C2A01081461).

\*\* 러시아리서치 센터 대표

주하는 토착소수민족들의 인구감소 및 민족문화의 해체나 소멸이 더욱 가속화 될 것으로 보인다.

▪ 주제어: 러시아, 북극개발, 소수민족, 텍스트마이닝, 뉴스분석

## I. 들어가는 말

러시아연방 내에 거주하는 소수민족들은 불안정한 정치·사회적 환경 속에서 살고 있다. 이는 한편으로는 이들의 권리를 보장해 주는 법적 규범이 아직까지 불완전한 탓도 있지만, 이들에 대한 법적 규범을 확립할 수 있는 정치·사회적 안정적인 토대가 마련되지 않았기 때문이다. 러시아연방정부의 소수민족정책은 주로 명목민족, 즉 민족공화국과의 관계에 집중되어 있다. 또한 최근 기후 온난화로 북극항로 상용화 및 북극 자원개발 가능성이 커지면서, 2000년대 급속한 경제성장을 발판으로 북극개발에 역량을 집중하게 되었다. 그러나 러시아의 북극개발은 북극토착소수민족들의 삶과 문화에 큰 영향을 주어 민족소멸의 위기로 확대될 수 있는 계기가 될 수도 있다.

러시아 정부의 북극에 대한 관심이 정책적인 기반을 갖추게 된 것은 2013년 2월 푸틴 대통령이 ‘러시아 연방 북극권 개발 전략’을 공포하고 ‘북극 정책 2020’을 기반으로 한 분야별 실천 과제를 제시함으로써, 러시아 정부의 북극 정책 추진 체계는 완성되게 된다(에너지경제연구원 2016, 17). 이후 2020년에 기존 북극정책 추진기한이 만료됨에 따라 “2035 러시아 북극정책 기본 원칙”이 마련되었다. 블라디미르 푸틴(Vladimir Putin) 러시아 대통령은 2020년에 ‘2035 러시아 북극 정책 기본원칙(Basic Principles of Russian Federation State Policy in the Arctic to 2035)’을 발표하였는데, 이 전략에서는 2020년부터 2035년까지 북극 온난화로 인한 자원개발 촉진과 북극 지역 인구 생활환경 개선 중심으로 계획이 수립되었다.

이 전략은 러시아 북극의 사회개발, 경제발전, 인프라 개발, 과학기술 개발, 환경보호, 국제협력 강화, 인구 및 영토 보호, 공공안전 확보, 군사안보

보장, 국경 보호로 나뉘어 있고, 장기적으로 북극항로 건설과 영토 보호를 위한 군사력 재건과 보존을 위한 계획을 제시하고 있다(경제·인문사회연구회 2021, 23). 2020년에 발표한 ‘2035 러시아 북극 정책 기본원칙’에서는 주요 정책 목표 중 하나가 사회·경제 개발 분야에서 북극권 거주민의 삶의 질과 복지를 보장하고 의료지원과 인재 양성을 하는 것이다. 이 역시 북극 지역 거주민 삶의 개선을 주요 과제로 삼고 있다.

북극 지역 주민이 안정적으로 거주할 수 있는 사회 환경을 구축하는 것에 정부의 초점이 맞춰지고 있으나 “2035 러시아 북극 정책 기본원칙”에서 북극에 거주하는 소수민족들에 대한 정책이 반영되지 않은 점은 소수민족에 대한 러시아 국민 및 정부의 의지가 반영된 것이라고 볼 수 있다.

러시아 국민 및 정부의 자국 거주 소수민족, 특히 북극에 거주하는 소수민족들에 대한 생각을 알기 위해서는 러시아에서 발행되는 뉴스를 분석하는 것이 중요하다. 또한 분석의 결과에 따라 북극 거주 소수민족에 대한 국내 연구의 방향성을 효율성 있도록 재조정할 필요가 발생할 수 있다.

매년 기하급수적으로 증가하는 뉴스 데이터는 매년 방대한 규모의 자료를 생산해낼 뿐 아니라 그 역사도 100년에 가까운 정도로 오래되었다는 점에서 빅데이터뿐 아니라 롱데이터(long data)로서 활용될 수 있다(김일환 2019, 43).

본 연구는 ‘타스 통신사’ 뉴스 텍스트의 데이터 분석을 통해 사건들 상호간의 관련성을 판단하여 네트워크를 형성함으로써 요약된 사건정보를 제공하고 이를 위해 텍스트 분석기법 및 연관분석 기법을 활용하여 정치사회적 이슈를 시스템적으로 분석한다.

## II. 선행연구 분석

국내에서 ‘뉴스 분석 빅데이터’를 분석 방법으로 사용한 대표적인 사례 논문들에는 “뉴스 기사의 빅데이터 분석 방법으로서 뉴스정보원 연결망 분석”(박대민 2013), “장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성: 100만 건 기사의 정보원과 주제로 본 신문 26년”(박대민 2016),

“뉴스 빅데이터를 활용한 평생교육 토픽 분석”(김태종·박상욱 2019), “서울의 다문화 공간 연구: 뉴스 빅데이터 분석 시스템 ‘빅카인즈’를 이용한 국내 언론의 외국인 마을 보도(1990-2016) 분석”(이은별 외 2017), “빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론보도 분석”(유예림 2017), “도서관에 대한 언론 보도 경향: 1990-2018 뉴스 빅데이터 분석”(한승희 2019) 등 수많은 연구 논문들이 있다. 이들 국내 뉴스에 관한 논문이 출현하는 데에는 한국언론재단의 텍스트 분석 시스템인 ‘빅카인즈’<sup>1)</sup>가 많은 도움을 주고 있기 때문이다. ‘빅카인즈’는 전국 언론 뉴스 데이터 베이스로 자체적으로 데이터 가공을 하여 연구자로 하여금 검색 및 분류, 데이터 시각화 등이 손쉽게 작성될 수 있도록 구성되어 있다.

그러나 텍스트 분석 시스템이 제공되지 않는 외국 뉴스 기사에 대한 연구는 영어권 등 일부 외국어권에 한정되고, 특히 러시아어로 된 뉴스 분석에 관한 연구는 거의 없는 실정이다. 국내에서 러시아 뉴스를 분석한 논문에는 “러시아 신문에 보도된 한반도 뉴스 특성에 대한 연구: 2002-2010년의 이즈베스티야에 보도된 한반도 관련 기사 내용분석”(RADJABOVA 2011), “R 텍스트 마이닝을 활용한 러시아 진보계열 미디어의 ‘크림반도 합병’ 관련 기사 분석: ‘RadioFreeEurope/RadioLiberty’, ‘Новая газета’의 2014-2015년, 2019년 기사 비교를 중심으로”(박진택 2019), “북극에 관한 러시아 언론분석 및 한국의 대응전략”(계용택 2015), “러시아 언론의 북극에 관한 보도내용 및 성향 분석: 타스통신사의 뉴스기사 텍스트를 중심으로”(계용택 2020) 등이 있다.

“러시아 신문에 보도된 한반도 뉴스 특성에 대한 연구”는 외국학자가 석사논문으로 러시아 이즈베스티야 신문 뉴스 텍스트를 분석한 경우다. “R 텍스트 마이닝을 활용한 러시아 진보계열 미디어~”는 빅데이터 분석 방법을 사용하였지만 분석 데이터로 사용된 이즈베스티야 기사 건수가 적고 러시아어가 아닌 영문 텍스트를 사용한 것이 한계점이다.

“북극에 관한 러시아 언론분석 및 한국의 대응전략”은 여러 러시아 매체에서 발행된 러시아어 뉴스 텍스트를 ‘빅데이터 분석 방법’을 사용한 연구이

1) 신문, 방송 등을 포함한 국내 최대의 기사 데이터베이스로 빅데이터 분석 기술로 뉴스 분석 서비스를 제공한다.

지만 연어 분석 및 단어 상관관계 분석 등이 충분하지 않다. “러시아 언론의 북극에 관한 보도내용 및 성향 분석: 타스통신사의 뉴스기사 텍스트를 중심으로”는 국내에서는 최초로 러시아어 뉴스 분석에서 연어 및 상관계수 등이 포함된 ‘빅데이터 분석 방법’을 충분히 활용한 연구논문이라 볼 수 있다.

### Ⅲ. 자료수집

본 연구는 2004년 11월 23일부터 2022년 1월 5일까지 약 17년간의 ‘타스통신사<sup>2)</sup>의 뉴스 기사 텍스트 가운데 ‘북극, 민족(арктика, народ)’으로 검색된 뉴스를 분석의 대상으로 삼았다. 논문의 분석에 사용된 전체기사는 2,401건이며 전체 단어수는 426,063개, 중복을 포함하지 않는 단어수는 31,848개이다.

### Ⅳ. 연구방법

#### 1. 텍스트 데이터 분석방법

데이터 정제 과정기법 중의 하나인 텍스트마이닝은 자연어처리(Natural language processing) 기술을 이용하여 비정형 혹은 반정형 텍스트 데이터로부터 유용한 정보를 추출 및 가공을 목적으로 하는 기술이다(전채남·서일원 2013). 텍스트마이닝 기술을 통해 방대한 텍스트 문치에서 유의미한 정보를 추출하고, 다른 정보와의 연계성을 파악하거나 텍스트가 가진 범주를 찾아내는 등 단순한 정보검색 그 이상의 결과를 얻어낼 수 있다(박만희 2016, 393). 형태소 분석은 텍스트 구성의 가장 최소단위인 형태소(단어)로 텍스트

---

2) 러시아 통신사 타스는 러시아의 국영 통신사이며 전신은 소련의 국영 통신사인 타스이다. 본사는 모스크바에 있으며 1992년 이타르타스라는 이름으로 설립했다. 2014년 9월 타스로 변경.

를 분석하는 가장 보편적인 방법이다. 이와 더불어 형태소 분석은 문법적 규칙 혹은 확률에 의한 품사 태깅(part of speech tagging), 개체명 인식(named entity recognition), 철자 교정, 단어 식별(tokenization) 기법 등으로 발전하여 분석효율을 높이고 있다.

동시출현 단어 분석은 단어들 사이의 의미상의 관계성을 파악하기 위해 일정한 문맥 내에서 두 단어가 동시 출현하는 빈도를 구한 후, 다양한 통계적 방법을 활용하여 유의미한 단어 쌍을 추출해낼 수 있다.

기본적으로 동시출현 단어 분석은 단어들 사이의 의미상의 관계성을 파악하기 위해 일정한 문맥 내에서 두 단어가 동시 출현하는 빈도를 구한 후, 다양한 통계적 방법을 활용하여 유의미한 단어 쌍을 추출해낼 수 있다. 즉 동시 출현한 단어 개수 및 관계를 파악하여 단어 간의 관계를 파악하고 키워드의 동시출현 패턴을 발견한다. 이는 뉴스 텍스트의 단어 간의 동시 출현 행렬을 생성하여 네트워크 형태로 표현하는 방법이다. 이를 통해 네트워크의 전체 형태와 개별 단어 간의 역할 설명을 가능하게 한다.

텍스트 데이터 분석에 주로 쓰이는 문자열 분석은 영어의 음운이나 한국어의 글자의 개수(n-gram)를 지정하여 전체 텍스트 코퍼스를 분석함으로써 해당 문자열이 나왔을 때 그다음에 어떤 글자가 나올지를 확률분포를 통해 예측하는 N-gram<sup>3)</sup> 알고리즘을 사용한다.

그밖에 출현단어 개수에 따라 단어의 크기가 변하는 그림으로 보여주는 워드클라우드 기법이 있다. 워드클라우드에는 문서에 사용된 단어의 빈도수에 따라 많이 나오는 단어는 크게 표시되기 때문에 한눈에 키워드를 파악할 수 있다.

머신러닝 등에서 사용되는 벡터 공간 모델은 살튼(Salton 1989) 등에 의해 1970년대에 주창되었고, 지금도 활발히 활용되는 방법이다. 벡터 공간 모델에서 문서(document)와 단어(term)를 각각 벡터(vector)와 차원(dimension)에 대응시켜 통계적 가중치를 구한다. TF-IDF(Salton & McGill 1983), 카이제곱 검정, 코사인 유사도 등을 통해 단어의 빈도수 분포에서 중요하지 않은 단어를 걸러내고, 문서의 유사도를 구한다. 이 모델을 적용한 기법으로 의미연결망 분석 및 토픽 모델링 방법이 있다. 의미연결망 분석은 키워드

3) n-gram은 n개의 연속적인 단어 나열을 의미한다. 갖고 있는 코퍼스에서 n개의 단어 문치 단위로 끊어서 이를 하나의 토큰으로 간주한다. n값이 1이면 유니그램(unigram), 2이면 바이그램(bigram), 3이면 트라이그램(trigram)이다.

동시출현 분석 기법에 기반하며, 단어 간의 의미 혹은 맥락상의 연결 관계를 정의하고, 해당 연결 관계를 시각화하거나 중요한 컨셉을 네트워크 속의 위상(centrality)에 따라 추출하는 방법이다(Wasserman & Faust 1994).

의미연결망 분석은, 단어출현 빈도수 및 하나의 문장에서 동시 출현하는 단어들의 관계로 텍스트의 의미화 패턴을 분석한다. 이때, 정보단위가 되는 단어나 구를, 각각의 노드(node)를 형성하는 개념으로 놓고 개념 간의 연결 상태를 링크(link)로 나타낸다. 여기서 링크로 드러나는 단어들의 공동출현 관계를 통해 의미를 해석한다(Wang & Rada 1998). 단어의 사용 빈도와 상호관계는 텍스트의 주된 상징성을 보여주고 단어들의 결합을 통해 문장 속에 숨어있는 의미를 보여주기도 한다.

출현하는 단어들의 상관분석(Correlation Analysis)은 두 연속 변수가 서로 관련이 있는지 검정하는 통계 분석 기법으로, 상관분석을 통해 도출되는 상관계수를 통해 두 변수의 관련성을 파악한다. 상관계수는 0-1 사이의 값을 지니고 1에 가까울수록 관련성이 크다는 것을 의미한다.

토픽 모델링은 구조화되지 않은 대량의 텍스트로부터 숨겨져 있는 주제 구조를 발견하고 카테고리화하기 위한 통계적 추론 알고리즘으로, LDA(Latent Dirichlet Allocation) 모델이 주로 활용되며 각각의 창발적인(emergent) 주제를 각 행에 배열되는 단어들의 확률분포를 통해 표현한다(Blei et al. 2003).

텍스트의 감성 분석은 감성 사전을 기반으로 분석 대상이 되는 전체 텍스트의 감성 비율을 정량화하는 기법이다. 통상적으로 긍정, 중립, 부정 등의 ‘평가어’ 분석이 감성분석으로 이해되나 이는 실제론 기쁨, 우울, 화남 등 심리학적인 감성 카테고리에 기반한 정량화의 한 특수한 사례이다.

그밖에 텍스트 분석에서 활용되는 머신러닝 기법들에는 SVM(Support Vector Machine),<sup>4)</sup> 의사결정트리(Decision Tree),<sup>5)</sup> 랜덤 포레스트(Random Forest),<sup>6)</sup> 딥 러닝(Deep Learning)<sup>7)</sup> 기법 등이 활용되고 있다.

- 
- 4) 서포트 벡터 머신은 인공지능의 기계학습 분야 중 하나로, 패턴인식, 자료분석을 위한 지도학습 모델로, SVM 알고리즘은 주어진 데이터 집합을 바탕으로 하여 새로운 데이터가 어느 카테고리에 속할 것인지 판단하는 비확률적 이진 선형 분류 모델을 만들게 된다.
  - 5) 의사결정트리는 일련의 분류 규칙을 통해 데이터를 분류(Classification)와 회귀(Regression) 모두 가능한 지도 학습 모델 중 하나로 결과 모델이 Tree 구조를 가지고 있다.
  - 6) 랜덤 포레스트는 대표적인 배깅(bagging) 방식 앙상블 알고리즘으로 여러 개의 트리 만들어 보팅으로 최종 클래스 값을 결정하는 알고리즘이다.

## 2. 본 연구에서 사용한 방법

본 연구는 뉴스기사 내용 분석을 위하여 동시 출현 단어 빈도수 분석 기법을 이용하여 키워드의 네트워크 구조와 바이그램 및 트라이그램으로 내용을 가시화하고, 이를 바탕으로 뉴스 기사 텍스트 전체의 내용과 구조를 파악할 것이다. 본 연구에서 수행한 주요 연구 절차와 핵심 연구내용 및 연구 방법론을 개략적으로 설명하면 다음과 같다.

본 연구에서는 텍스트마이닝을 위해 R의 TM 패키지를 이용하였다. 뉴스 기사 정제를 위해 텍스트의 공백제거, 마침표 제거, 불용어 처리, 소문자 변환, 단어 형태소 분석-변환 등을 실시한다.

단어들 및 문서 간의 관계를 파악하기 위해 단어문서행렬(Term document matrix)을 생성하는 기법을 사용하였다. 이와 더불어 텍스트마이닝 결과 도출 모델 및 시각화에 N-gram(바이그램, 트라이그램) 및 워드클라우드, 의미연결망 모델을 사용하였다.

사용된 주된 컴퓨터 프로그램으로는 R(Rstudio) 및 분석모듈, 파이썬(Python) 및 분석모듈, MS Window Server 및 MS SQL Server 등이 있으며 스크립트 코딩에는 Visual Studio를 사용하였다.

작업환경으로 웹 브라우저에서 파이썬 코드를 작성하고 실행까지 해볼 수 있는 아나콘다 주피터 노트북(Anaconda Jupyter Notebook)을 사용하였으며, 수집된 뉴스 기사를 발행날짜, 인터넷주소, 기사 내용 등의 레코드로 구성하여 문장 단위로 분리된 레코드가 있는 데이터베이스를 구축하였다.

## V. 연구내용 및 결과

### 1. 뉴스기사 성향 분석

‘북극, 민족(арктика, народ)’으로 검색된 전체 기사건수는 모두 2,401건

- 7) 여러 비선형 변환기법의 조합을 통해 높은 수준의 추상화를 시도하는 기계 학습 알고리즘의 집합으로, 사람의 사고방식을 컴퓨터에게 가르치는 기계학습의 한 분야이다.

이다. 이들 뉴스를 타스 통신사의 뉴스 카테고리별로 분류하면 다음과 같다 (숫자는 기사 건수).

[사회 689, 국내 360, 경제 314, 정치 149, 국제-파노라마 134, 과학 113, 뉴스-파트너들 104, 우랄-뉴스 83, 문화 71, 시베리아-뉴스 62, 인터뷰 52, 북극-오늘 43, spb-뉴스 33, 사건사고 32, 아르히브 28, 프레스 15, 인포메이션 15, 국가-프로젝트 15, 군대-군산복합체 13, 스포츠 11, 오피니언 7, 모스크바 6, plus-one 6, 우주 6, 중소기업 6, 모스크바-지역 4, 북카프카즈 4, 지역-공무원 3, 아르한겔스크 주 3, 사람들-물건 3, 보도 자료 3, 사이언스 2, 무르만스크 주 2, 온라인-컨퍼런스 1, 고급-공무원 1, 당일 투표 1, vef-2017 1, 러시아 관광 1].

검색된 뉴스의 대부분은 사회, 국내, 경제, 정치, 국제-파노라마, 과학, 뉴스-파트너, 우랄-뉴스, 문화, 시베리아-뉴스 카테고리 등 다양한 분야에 속해 있기 때문에 편향되지 않은 다양한 내용의 뉴스가 포함되었다고 볼 수 있다. <표 1>에 제시된 뉴스 카테고리별 뉴스 제목들을 보면 다양한 주제를 포함한 내용을 볼 수 있는데 이는 키워드 검색이 편중되지 않았음을 제시하고, 러시아 정부 및 국민의 성향이 충분히 포함되어 있다고 볼 수 있다.

<표 1> 검색된 뉴스의 URL주소 및 뉴스기사 제목

URL주소	뉴스기사 제목
<a href="https://tass.ru/obschestvo/10022403">https://tass.ru/obschestvo/10022403</a> (사회)	Комитет Думы одобрил законопроект об увеличении сроков перерегистрации охотничьего оружия
<a href="https://tass.ru/obschestvo/10094463">https://tass.ru/obschestvo/10094463</a> (사회)	Глава Якутии предложил провести в 2023 году всемирный мерзлотный саммит в Якутске
<a href="https://tass.ru/v-strane/10013483">https://tass.ru/v-strane/10013483</a> (국내)	Быть чамалинцем: почему один из народов Дагестана оказался самым малочисленным в России
<a href="https://tass.ru/v-strane/10214079">https://tass.ru/v-strane/10214079</a> (국내)	Представители КМНС из ЯНАО получают целевое образование в вузе Санкт-Петербурга
<a href="https://tass.ru/ekonomika/10024653">https://tass.ru/ekonomika/10024653</a> (경제)	Эксперты: окрестности городов и берега рек перспективны для выдачи “арктического гектара”
<a href="https://tass.ru/ekonomika/10095025">https://tass.ru/ekonomika/10095025</a> (경제)	РФ в рамках председательства в Арктическом совете усилит социально-экономическую повестку
<a href="https://tass.ru/politika/1356697">https://tass.ru/politika/1356697</a> (정치)	Патрушев указал на рост экстремизма в СЗФО

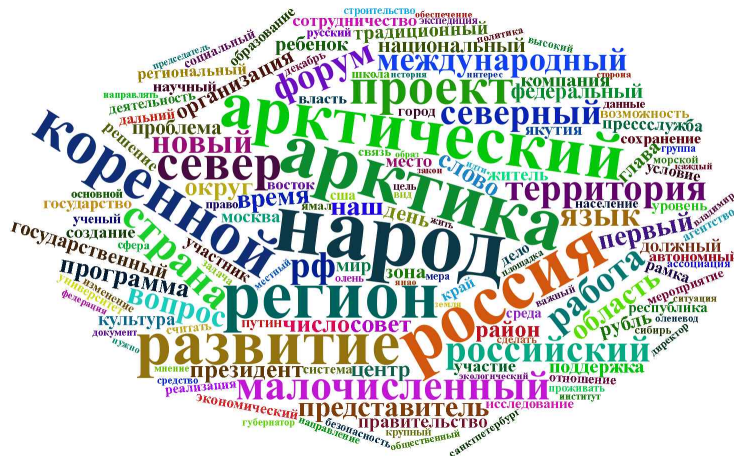
URL주소	뉴스기사 제목
<a href="https://tass.ru/politika/12260005">https://tass.ru/politika/12260005</a> (정치)	Почти 500 представителей коренных малочисленных народов Севера досрочно проголосовали
<a href="https://tass.ru/mezhdunarodnaya-panorama/10075023">https://tass.ru/mezhdunarodnaya-panorama/10075023</a> (국제)	Эволюция протеста: почему в Таиланде хотят поменять конституцию
<a href="https://tass.ru/mezhdunarodnaya-panorama/13074315">https://tass.ru/mezhdunarodnaya-panorama/13074315</a> (국제)	Правящая партия Гондураса признала поражение своего кандидата на президентских выборах
<a href="https://tass.ru/nauka/10493481">https://tass.ru/nauka/10493481</a> (과학)	Семена борщевика занесли в Арктику
<a href="https://tass.ru/nauka/12032535">https://tass.ru/nauka/12032535</a> (과학)	Большая норильская экспедиция проведет экспертизу Норило-Пясинской водной системы
<a href="https://tass.ru/novosti-partnerov/10126721">https://tass.ru/novosti-partnerov/10126721</a> (뉴스-파트너)	Фонд развития Югры поддержал продвижение производителей ХМАО в федеральных торговых сетях
<a href="https://tass.ru/novosti-partnerov/12688497">https://tass.ru/novosti-partnerov/12688497</a> (뉴스-파트너)	Наталья Фиголь: молодежь готова жить в Арктике, если видит возможности самореализации
<a href="https://tass.ru/ural-news/10185129">https://tass.ru/ural-news/10185129</a> (우랄)	Власти ЯНАО выделили более 47 млн рублей на поддержку коренных народов Севера в пандемию
<a href="https://tass.ru/ural-news/11080535">https://tass.ru/ural-news/11080535</a> (우랄)	Югра приобретет второй медицинский теплоход для обследования жителей отдаленных поселков
<a href="https://tass.ru/kultura/10054143">https://tass.ru/kultura/10054143</a> (문화)	В эвенкийском селе в Амурской области построят этнокультурный центр
<a href="https://tass.ru/kultura/13073937">https://tass.ru/kultura/13073937</a> (문화)	Сокуров представит в Турине мировую премьеру восстановленной версии фильма “Молох”
<a href="https://tass.ru/sibir-news/10087727">https://tass.ru/sibir-news/10087727</a> (시베리아)	Турпоток на Тай мьр в 2020 году превысил показатель и прошлого года
<a href="https://tass.ru/sibir-news/11624625">https://tass.ru/sibir-news/11624625</a> (시베리아)	“Дочка” “Норникеля” дополнительно выплатит почти 16 млн рублей представителям КМНС
<a href="https://tass.ru/interviews/10512039">https://tass.ru/interviews/10512039</a> (인터뷰)	Министр по развитию Дальнего Востока и Арктики: лед будет проломлен
<a href="https://tass.ru/interviews/5554565">https://tass.ru/interviews/5554565</a> (인터뷰)	Глава ФАДН России: прививку от национализма нужно делать еще в детском саду
ФАДН = Федеральное агентство по делам национальностей КМНС = Коренные малочисленные народы Севера	

## 2. 뉴스기사 단어 빈도수 분석

기사의 뉴스 기사 본문 텍스트에 나타난 단어의 빈도수 분석 결과, 빈도수가 높은 상위 50개 단어 및 빈도수는 다음과 같다.

[민족 5,643, 북극 4,898, 러시아 4,597, 지역 4,377, 북극의 4,328, 발전 4,062, 토착의 4,025, 프로젝트 3,289, 북방 3,215, 국가 2,727, 소수의 2,711, 러시아의 2,445, 영토 2,422, 일(업무) 2,371, 러연방 2,366, 북방의 2,321, 포럼 2,280, 언어 2,196, 새로운 2,171, 국제적인 2,144, 문제 2,043, 주 1,839, 대표자 1,823, 우리 1,764, 첫째의 1,748, 시간 1,683, 프로그램 1,672, 단어 1,666, 관구 1,656, 위원회 1,613, 숫자 1,608, 대통령 1,576].<sup>8)</sup>

<그림 1> 검색된 단어들로 구성된 워드클라우드



8) народ 5,643, арктика 4,898, россия 4,597, регион 4,377, арктический 4,328, развитие 4,062, коренной 4,025, проект 3,289, север 3,215, страна 2,727, малочисленный 2,711, российский 2,445, территория 2,422, работа 2,371, рф 2,366, северный 2,321, форум 2,280, язык 2,196, новый 2,171, международный 2,144, вопрос 2,043, область 1,839, представитель 1,823, наш 1,764, первый 1,748, время 1,683, программа 1,672, слово 1,666, округ 1,656, совет 1,613, число 1,608, президент 1,576, район 1,515, центр 1,424, организация 1,407, день 1,397, федеральный 1,358, национальный 1,351, проблема 1,306, глава 1,298, ребенок 1,295, зона 1,292, рубль 1,280, поддержка 1,263, традиционный 1,245, компания 1,240, должный 1,231, правительство 1,230, мир 1,218, культура 1,177, место 1,175, москва 1,157



검색어인 민족(народ)을 중심으로는 ‘토착의, 언어, 북부, 소수의’ 단어들 이 굵은 선으로 연결되어 있다. 이는 ‘토착 소수 민족 언어’의 상관도가 높을 것을 보여주는 것이다.

#### 4. 뉴스기사 엔그램(n-gram) 단어 간 네트워크 분석

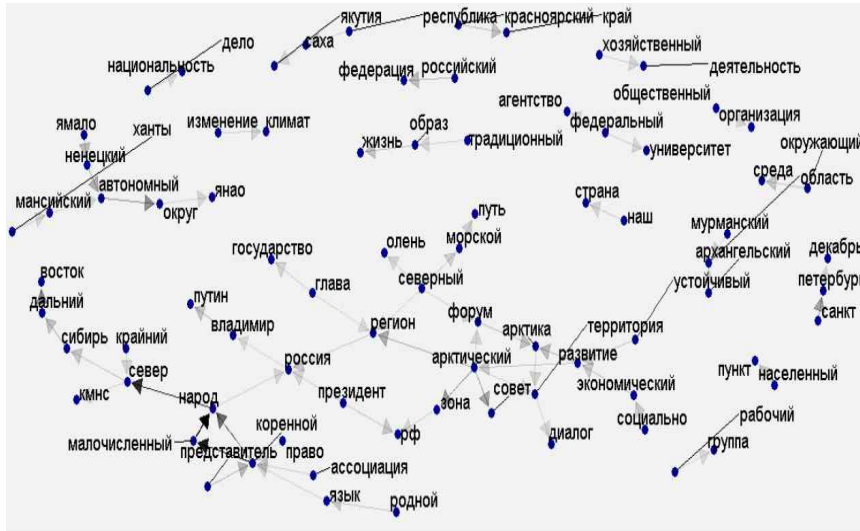
##### (1) 바이그램 분석

텍스트의 의미를 제대로 조합하기 위해서는 개별 단어보다 크고 문장보다 작은 단위가 중요하며 이를 위해 단어의 연쇄 추출이 가능한 엔그램(n-gram) 방법을 사용한다. 언어 분석을 통해 나타나는 단어들의 연결망에서는 단어들 사이의 관계가 나타나는데 가장 많이 사용하는 2단어 언어 분석(bi-gram)은 이웃하는 2개 단위로 묶어 분석하며, 1개 단어만으로는 이웃하는 단어들의 의미가 상실되는 점을 보완하는 용도로 사용한다. 2단어 언어분석은 연산 결과표만으로는 의미 해석이 힘들어 언어 분석 관계를 시각화하여 관계를 살펴보는 것이 요구된다.

<그림 3>은 전체 기사 본문에서 추출한 바이그램 단어 쌍 및 단어 쌍 간의 네트워크를 표현한 것이다. 바이그램은 한 문장에서 나타나는 2개의 연속되는 단어 가운데 빈도수 높은 것들을, 의미가 연결되는 단어 쌍으로 보여 줄 수 있다. 그러므로 바이그램은 구체적인 의미를 파악하는 데 높은 효용성을 가진다. 뉴스 기사 본문 전체를 표현한 바이그램은 뉴스 기사 내용을 대표되는 몇 개의 키워드 그룹으로 나누어 질 수 있게 한다. 바이그램 단어 들에는 소수 민족, 북방 민족, 토착 민족, 자치 관구, 극동, 북극 존, 쌍-페테르부르그, 북극 위원회, 북극 지역, 네네쯔 자치관구, 토착민족 대표자, 야말-네네쯔키이, 러시아 연방, 북극 개발, 북극 포럼, 크라스노야르 지방, 블라디미르 푸틴, 거주 형태, 북부 시베리아, 러시아연방 대통령 등이 있다.<sup>9)</sup>

9) малочисленный народ 2,593, коренной малочисленный 2,248, народ север 2,044, коренной народ 1,385, автономный округ 952, дальний восток 900, арктический зона 819, Санкт-Петербург 753, арктический совет 750, арктический регион 643, ненецкий автономный 590, представитель коренной 535, Ямало-Ненецкий 471, российский федерация 467, развитие арктика 462,

<그림 3> 뉴스기사 텍스트 바이그램



<그림 3>을 보면 단어연결 고리가 몇 가지 그룹으로 나누어져 있음을 보여주고 있다. 여기에는 ‘북극지역개발’, ‘북방 소수토착민족’, ‘북방 해운 항로’, ‘야말로-네네쯔 한티-만시 자치관구’, ‘무르만스크 아르한겔스크’ 그룹으로 나누어져 있다. 이 중 가장 많은 단어가 연결된 그룹으로는 ‘북극지역개발’ 및 ‘북방 소수토착민족’ 그룹이다.

forum arktiki 416, sibir dallyniy 409, krasnoyarskiy kraj 407, vladimir putin 405, obraz zhizn 400, sever sibir 383, prezident rf 350, okruzhayushiy sreda 343, ekonomicheskiy razvitiye 337, rodnoy yazyk 332, razvitiye arkticheskiy 331, region rossiya 331, ustay chivyy razvitiye 330, sotsialno ekonomicheskiy 314, naselennyy punkt 298, mурманский область 277, severnyy morskoy puty 277, morskoy puty 268, assotsiatsiya korennoy 253, arktika territoriya 250, arkhangel'skiy область 247, krajniy sever 242, izmeneniye klimata 233, zona rf 232, yazyk korennoy 230, prezident rossiya 229, nash strana 218, glavа gosudarstvo 214, narod rossiya 214, xanty mansiyskiy 213, territoriya dialog 210, arkticheskiy forum 205, severnyy forum 203, hozyaystvennyy deyatelnosty 202, okrug yanao 198, sever kmns 198, federalnyy universitet 198, arkticheskiy territoriya 197, traditsionnyy obraz 194, respublika саха 189, severnyy oleny 187, саха якутия 185, federalnyy agentstvo 184, mansiyskiy avtonomnyy 179, delo natsionalnosty 177

## (2) 트라이그램 분석

다음은 200회 이상 출현한 트라이그램 및 빈도수를 나타낸 것이다.

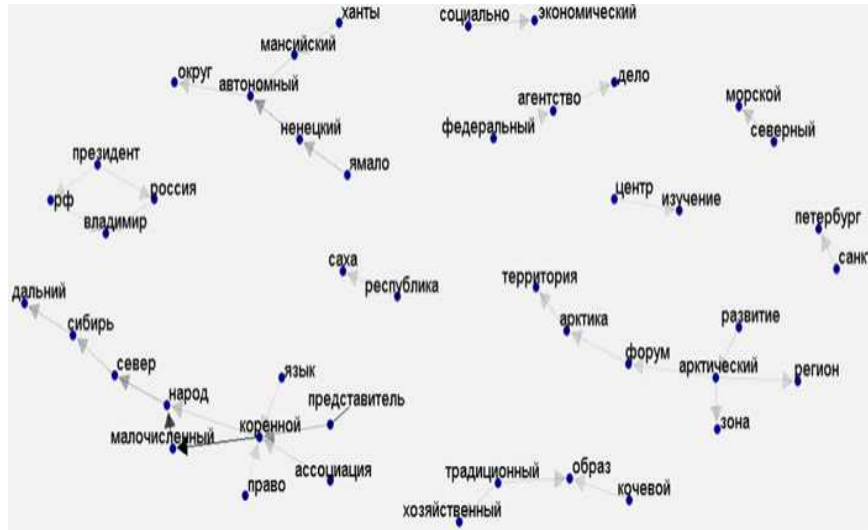
여기에는 [토착 소수민족 2,240, 북방 소수민족 1,526, 네네쯔 자치관구 586, 시베리아 극동 409, 야말로 네네쯔 자치 400, 북방 시베리아 377, 토착 민족 대표 366, 시베리아 북방 민족 347, 북방소수민족 262, 북방 바닷길 261, 토착 소수 협회 240, 리연방 북극 존 231, 사회 경제 발전 206, 북방 영토 대화 203] 등이 있다.<sup>10)</sup>

트라이그램(trigram)은 하나의 문장에서 3개의 연속된 단어들의 묶음으로 보여지기 때문에 문장의 구체적인 의미를 파악하는 데 바이그램보다 효용성이 높다. 뉴스 기사 본문 전체를 표현한 트라이그램은 빈도수가 높은 키워드 그룹으로 나누어져 뉴스 기사를 구성하고 있는데 여기에는 ‘토착 북방 소수 민족’, ‘야말로 네네쯔 자치관구’, ‘북방 시베리아 극동’, ‘북방 바닷길’, ‘리연방 북극 지역’, ‘사회 경제 발전’ 그룹 등이 있다. <그림 4>에서 핵심단어 연결그룹으로 ‘소수 토착 민족의 언어 및 권리’, ‘러시아 연방 대통령’, ‘야말로-네네쯔, 한티-만시 자치관구’, ‘연방 협회 업무’, ‘북극지역 발전 포럼’, ‘민

10) коренной малочисленный народ 2,240, малочисленный народ север 1,526, ненецкий автономный округ 586, сибирь дальний восток 409, ямало ненецкий автономный 400, север сибирь дальний 377, представитель коренной малочисленный 366, народ север сибирь 347, коренной народ север 262, северный морской путь 261, ассоциация коренной малочисленный 240, арктический зона рф 231, социально экономический развитие 206, арктика территория диалог 203, автономный округ янао 197, народ север кмн с 196, форум арктика территория 192, традиционный образ жизнь 191, развитие арктический зона 179, ханты мансийский автономный 179, мансийский автономный округ 178, республика саха якутия 178, санкт петербург декабрь 161, представитель коренной народ 149, россия владимир путин 149, президент россия владимир 148, федеральный агентство дело 146, агентство дело национальность 138, арктический регион россия 135, рф владимир путин 122, президент рф владимир 121, язык коренной народ 120, традиционный хозяйственный деятельность 112, арктический форум арктика 111, кочевой образ жизнь 105, право коренной народ 104, центр изучение арктика 103, язык коренной малочисленный 100, развитие дальний восток 99, дальний восток арктика 96, коренная малочисленный народ 96, научный центр изучение 94

족의 전통 경제 및 유목 생활'이 표시되고 있다.

<그림 4> 뉴스기사 텍스트 트라이그램



### 5. 북극토착소수민족별 상관도 분석

전체 텍스트에서 ‘민족(народ)’ 단어와 상관관계가 높은 민족들을 상관계수가 높은 순으로 보면 한티족(ханты 0.22), 에벤족(эвен 0.17), 에벤키족(эвенк 0.14), 만시족(манси 0.14), 셀쿠프족(селькуп 0.14), 나가산족(нга сан 0.12), चुкчи족(Чукчи 0.12), 돌간족(долган 0.11), 사미족(саам 0.05) 등이 있다. 이것은 러시아 국민들에게 있어 북극토착소수민족의 대표로 한티족 및 에벤족, 에벤키족을 연상하게 된다는 것을 의미한다.

뉴스기사에서 북극토착소수민족에 연관된 요소로 민족의 전통, 언어, 생활 모습 등의 민족자체 특성 요소뿐만 아니라 북방토착소수민족협회 및 민족문제 관련 연방 위원회 등 공공-정부기관 등이 등장한다. 이것은 민족문제에서 자치단체 또는 공공적인 차원에서의 도움 및 관리가 필요하다고 보는 것을 의미한다.

<표 2> 북극토착소수민족별 상관도가 높은 단어들

키워드	상관도 높은 단어
민족 (народ) <sup>11)</sup>	토착의, 소수의, 북부, 전통의, 종족의, 민족성, 언어, 친족의, 보존, 대표(자), 협회, 문화, 연방민족부, 북방토착소수민족, 모습, 거주, 언어의, 전통, 태고의, 바리노프(북방토착소수민족협회장), 교과서, 가르치는 것, 인종의, 레드코프(연방위원회 회원), 업무, 네네쯔, 보스톡(восток), 시베리아(сибирь), 한티족, 종족, 교육
한티족 <sup>12)</sup>	네네쯔족, 셀쿠프족, 심볼, 오브스키(обский) 가스화학 콤플렉스, 유그라족, 만시족, 민족, 토착의, (각 청의)국, 유고르스크의, 자치(주)의, 북방토착소수민족, 소수(민족)의, 한티만시스크, 북방, 습관의, 의식(儀式), 노래, 야말로-네네쯔 자치구, 한티만시스크의, 한티-만시 자치구, 문학의, 구비문학, 명절, 모습, 명료한, 흑사병(페지)
에벤키족 <sup>13)</sup>	에벤키의, 아무르의, 타이가, 생활양식, 언어학자, 지인(교제), 언어, 돌간족, 친절함, 기둥 모양의 물건, 예카테리나, 느가나산족, 착용자(보유자), 장식적인, 마을, 사진, 문자(문헌), 자바이칼의, 치밀한, 종족의, 크라스노야르스크, 희박한, 일상의, 언어의, 방언(사투리), 대학학과, 순록썰매, 에벤키 자치구, 텍스트
사미족 <sup>14)</sup>	사미족의, 콜라(반도)의, 무르만스크, 무르만스크의, 지방(끝)의, 묘사, 주의(областно), 반도, 마을, 대표자, 박람회, 자작농, 눈의(снежный), 포포브, 고고학자, 장식(물), 거주시설, 박물관, 나호드카, 희망, 국회, 축전(의식), 대회, 주(область), 수속(법률화), 생활양식, 노르웨이, 물건, 제작자, 희망, 기록, 의식(儀式), 민속학의
돌간족 <sup>15)</sup>	돌간족의, 느가산족, 볼로찬카 마을, 타이미르, 창조, 촌락, 에벤키족, 치밀한, 무용, 네네쯔족, 포포브, 크라스노야르스크의, 자주적인, 의상, 가죽, 전통관례, 셀쿠프족, 노래, 문자(문헌), 두딘카, 에벤족, 섬리, 인종, 민족, 친족, 무상원조, 제조, 도구(악기), 인종의, 초중고생, 집(가정용)의, 크라스노야르스크
만시족 <sup>16)</sup>	스베르들로프의, 우랄의, 구매, 지위(상태), 비싼, 잡지, 토착의, 선택, 신용, 한티족, 행복, 테마, 문명(문화), 기숙사제 학교, 대학학과, 의미, 보통의, 한티족, 요구(수요), 전화, 유입(출입구), 출생, 고용(작업), 움직임, 사회(집단)
셀쿠프족 <sup>17)</sup>	네네쯔족, 한티족, 야말로-네네쯔 자치구, 북방토착소수민족, 모습, 언어학자, 토착의, 야말로네네쯔의, 프리우랄의, 대표자, 북방인들, 느가산족, 야말의, 타쭝스키(라이온), 언어, 전러시아의, 인구, 숫자, 문자(문헌), 나이 어린, 야말, 인종의, 지역연구, 민족, 공모 담당실, 돌간족, 인터넷, 푸로브스키이(라이온), 소수의
나가산족 <sup>18)</sup>	타이미르, 돌간족, 볼로찬카 마을, 섬리, 에벤족, 문자(문헌), 돌간족의, 마을, 네네쯔족, 종족, 두딘카, 크라스노야르스크의, 툰드라, 조상(선조), 인종의, 셀쿠프족, 툰드라의, 지방(끝)의, 전통, 크라스노야르스크, 스베틀라나, 유목의, 춤(천막), 에벤키족, 마을, 관습, 세기, 니켈
에벤족 <sup>19)</sup>	축치족, 하바로프의, 정체성, 종족의, 마가단의, 민족, 에벤키족, 지방(край), 교과서, 보조금, 시설, 언어학자, 토착의, 돌간족의, 소멸, 언어, 보스톡(восток), 프리모르스키의, 하바로프스크, 희망하는, 상실, 방언, 가르치는 것, 교수, 세베르보스토치니이, 연주, 소수의, 북방, 학습의, 유전학의

키워드	상관도 높은 단어
축치족 <sup>20)</sup>	추코트의, 추코트카, 에벤족, 조상(선조), 미하일, 정체성, 지향(의지), 상실, 음악의, 마을, 교역(거래), 희망하는, 아나디리(지명), 주거지, 전통, 민족, 종족의, 확산하는, 선명한, 거주자

- 11) 상관계수 коренной 0.82, малочисленный 0.69, север 0.45, традиционный 0.40, родной 0.39, национальность 0.39, язык 0.35, родные 0.31, сохранение 0.31, представитель 0.30, ассоциация 0.29, культура 0.28, фадна 0.28, кмнс 0.27, образ 0.26, проживание 0.26, языковой 0.26, традиция 0.25, исконный 0.24, баринов 0.24, учебник 0.24, преподавание 0.24, этнический 0.24, ледков 0.23, дело 0.22, ненец 0.22, восток 0.22, сибирь 0.22, ханты 0.22, этнос 0.22, владеть 0.22, коренная 0.21, образование 0.20, реестр 0.20
- 12) 상관계수 ненец 0.39, селькуп 0.39, символ 0.30, обский 0.26, югра 0.25, манси 0.23, народ 0.22, коренной 0.22, департамент 0.22, югорский 0.22, автономный 0.21, кмнс 0.20, малочисленный 0.19, хантымансий ск 0.19, вести 0.18, север 0.18, привычный 0.18, обряд 0.18, песня 0.18, янао 0.17, хантымансий ский 0.17, хмао 0.17, литературный 0.17, фольклор 0.17, праздник 0.16, образ 0.16, коренная 0.16, понятный 0.16, чума 0.15, этнограф 0.15, многочисленный 0.15, игра 0.15, округ 0.14, северянин 0.14
- 13) 상관계수 эвенка 0.55, эвенкий ский 0.45, амурский 0.25, тай га 0.24, быг 0.22, лингвист 0.21, знакомство 0.20, язык 0.18, садиться 0.17, долган 0.17, добрость 0.17, рубрика 0.17, екатерина 0.16, эвен 0.16, исчезать 0.15, нганасан 0.15, носитель 0.15, декоративный 0.15, село 0.14, учить 0.14, познакомиться 0.14, заведующая 0.14, словарь 0.14, научиться 0.13, письменность 0.13, забайкальский 0.13, компактный 0.13, родной 0.12, красноярск 0.12, зафиксировать 0.12, редкий 0.12, повседневный 0.12, языковой 0.12, диалект 0.12, жить 0.11, издавать 0.11, кафедра 0.11, нарты 0.10, эвенкия 0.10, текст 0.10, причем 0.10
- 14) 상관계수 саамский 0.77, кольский 0.52, мурманск 0.30, мурманский 0.27, краеведческий 0.25, изображение 0.22, областной 0.21, полуостров 0.21, деревня 0.18, делегат 0.18, экспозиция 0.17, хозяин 0.16, снежный 0.15, попов 0.15, археолог 0.15, украшение 0.15, жилище 0.14, музей 0.13, насчитывать 0.13, находка 0.13, хх 0.13, желание 0.12, парламент 0.12, празднование 0.11, съезд 0.11, область 0.10, оформление 0.10, быг 0.10, норвегия 0.10, вещь 0.10, создатель 0.10, надежда 0.10, рекорд 0.10, обряд 0.10, фольклорный 0.10
- 15) 상관계수 долганский 0.48, нганасан 0.42, волочанка 0.37, тай мырский 0.34, тай мыр 0.32, денис 0.29, энца 0.26, творчество 0.25, поселок 0.21, эвенка 0.21, компактный 0.18, танец 0.18, ненец 0.17, эвенк 0.17, попов 0.16, красноярский 0.16, самобытный 0.16, одежда 0.15, шкура 0.15, традиция 0.14, селькуп 0.14, песня 0.14, письменность 0.13, дудинка 0.13, эвен 0.13, попадать 0.12, промысл 0.12, этнос 0.12, жить 0.11, изготавливать 0.11, народ 0.11, родственник 0.11, грант 0.11, изготовление 0.11, инструмент 0.11, этнический 0.11, школьник 0.11, возродить 0.11, увидеть 0.10, домашний 0.10, красноярск 0.10, вьигрват

또한 토착소수민족에 따라 뉴스기사의 시각은 다양한 것으로 나타난다. 북극토착 소수민족으로 대표되는 한티족의 경우 ‘오브스키 가스화학 콤플렉스’ 및 한티 만시 자치구 관련 정치-경제적 실질적인 삶의 모습을 비중 있게 나타내는 동시에, 한티족의 전통, 의식(儀式), 노래, 구비문학, 명절 등과 같

ь 0.10

- 16) 상관계수 свердловский 0.70, уральский 0.35, пусть 0.31, сильно 0.30, покупка 0.30, статус 0.29, дорогой 0.29, интересоваться 0.27, делать 0.26, журнал 0.25, коренная 0.24, выбор 0.23, кредит 0.23, ханты 0.23, благо 0.22, тема 0.21, цивилизация 0.20, жить 0.19, интернат 0.19, факультет 0.19, значимость 0.19, обычный 0.18, хант 0.18, спрос 0.18, обязательно 0.18, соответственно 0.18, телефон 0.17, поступление 0.17, закрытый 0.17, род 0.16, занятость 0.16, динамика 0.16, общество 0.16
- 17) 상관계수 ненец 0.49, ханты 0.39, янао 0.33, хант 0.29, вести 0.28, кмнс 0.26, образ 0.25, лингвист 0.21, коренной 0.20, ямалоненецкий 0.20, приуральский 0.20, представитель 0.19, зафиксировать 0.19, северянин 0.18, увеличиваться 0.18, нганасан 0.18, ямальский 0.17, тазовский 0.17, энца 0.17, язык 0.16, всероссийский 0.16, население 0.16, численность 0.16, письменность 0.16, младший 0.15, ямал 0.15, этнический 0.15, регионоведение 0.15, народ 0.14, прессслужба 0.14, долган 0.14, интернат 0.14, пуровский 0.14, малочисленный 0.13
- 18) 상관계수 тай мьр 0.47, тай мьрский 0.45, долган 0.42, энца 0.40, волочанка 0.26, промысл 0.26, эвенка 0.26, письменность 0.25, долганский 0.24, поселок 0.23, ненец 0.23, род 0.21, дудинка 0.21, красноярский 0.19, тундра 0.18, предок 0.18, этнический 0.18, селькуп 0.18, разговаривать 0.18, тундровый 0.17, краеведческий 0.16, когда-то 0.16, традиция 0.16, денис 0.16, красноярск 0.16, светлана 0.16, кочевье 0.16, чум 0.15, утрачивать 0.15, эвенк 0.15, древний 0.14, обычай 0.14, передавать 0.13, век 0.13, никель 0.13, зафиксировать 0.13
- 19) 상관계수 чукча 0.28, хабаровский 0.27, идентичность 0.26, эвенка 0.21, родной 0.20, магаданский 0.20, народ 0.17, эвенк 0.16, желать 0.15, край 0.14, учебник 0.14, пособие 0.14, заведение 0.14, владеть 0.14, лингвист 0.14, коренной 0.13, долган 0.13, исчезновение 0.13, язык 0.12, восток 0.12, приморский 0.12, гарантировать 0.12, выявлять 0.12, хабаровск 0.12, желающий 0.12, утрата 0.12, диалект 0.12, преподавание 0.11, обуславливать 0.11, профессор 0.11, остро 0.11, северовосточный 0.11, игра 0.11, малочисленный 0.10, север 0.10, учебный 0.10, генетический 0.10
- 20) 상관계수 чукотский 0.56, чукотка 0.30, возрождать 0.29, эвен 0.28, родиться 0.28, бывший 0.27, предок 0.25, михаил 0.24, идентичность 0.23, побеждать 0.20, стараться 0.19, стремление 0.19, утрата 0.18, музыкальный 0.15, село 0.15, обучать 0.15, промысел 0.15, желающий 0.15, анадьрь 0.14, постепенно 0.14, жить 0.13, крепкий 0.13, жилище 0.13, традиция 0.13, бросать 0.13, закупать 0.13, уверять 0.13, самостоятельно 0.13, придумывать 0.13, народ 0.12, родной 0.12, уверенный 0.12, яркий 0.12, житель 0.12

은 민족적 요소를 중요하게 다루고 있다. 이에 반하여 돌간족의 경우 무용, 의상, 가죽, 전통관례, 노래, 악기, 문자(문헌), 거주지 등 주로 민족적 요소만을 강조하고 있다. 특히 축치족의 경우 전통, 음악 등의 민족적 특성과 더불어 민족의 고유성 및 지향점, 의지 등이 포함된 민족 정체성이 부각되었다. 만시족의 경우 민족적 요소보다 거주민으로서의 생활과 관계되는 직업활동, 사회활동, 교육과정 등에 중점을 두는 경향을 보이고 있다.

## VI. 결론

‘타스 통신사’ 뉴스 기사 텍스트 본문 분석에서 단어 빈도수(워드클라우드 제시) 분석, 단어 간의 네트워크 분석, 상관분석(Correlation Analysis)을 실시하였다. 엔그램(n-gram) 분석에서는 바이그램 분석, 및 트라이그램 분석을 통하여 보다 세밀한 연구를 시도하였다. 이러한 연구에서 도출된 키워드를 가지고 기사 내용의 본질적인 의도를 추론하였다. 검색된 기사들에서 출현빈도수가 높은 단어로는 [민족 5,643, 북극 4,898, 러시아 4,597, 지역 4,377, 북극의 4,328, 발전 4,062, 토착의 4,025, 프로젝트 3,289, 북방 3,215, 국가 2,727, 소수의 2,711, 러시아의 2,445, 영토 2,422, 일(업무) 2,371, 러연방 2,366] 등이 있다. 대부분의 단어들은 러시아의 북극개발 프로젝트와 관련된 단어들이다.

이에 반해 북극에 사는 토착민족들 관련 단어들의 출현 빈도수는 상대적으로 매우 적은 편이다. 대표적인 민족과 빈도수를 보면 [한티족 520, 에벤키족 332, 사미족 233, 돌간족 115, 만시족 112, 셀쿠프족 94, 벵스족 69, 나가산족 63, 유카기리족 63, 에벤족 51, 축치족 48, 카랴족 31, 알레우트족 30, 케트족 27, 우데케이족 26, 에스키모족 21, 니브호족 20, 이텔멘족 17, 울리치족 14, 오로치족 10]으로 나타난다. 그밖에 소수민족과 관련된 음악, 무용, 관습, 종교, 놀이, 거주지, 문자, 역사, 의복, 명절, 축제, 전통, 정체성, 음식 등의 단어가 상대적으로 매우 낮은 출현빈도를 보여준다.

바이그램 및 트라이그램, 상관관계 분석에서도 북극토착소수민족 단어들의 상관도는 북극개발 프로젝트에 비하면 매우 낮은 수준이라 볼 수 있다.

이러한 연구결과로 볼 때 검색된 ‘타스 통신사’의 전체 기사의 핵심 키워드는 ‘러시아 북극지역 사회경제적 개발’ 및 ‘북극지역 소수민족 언어 권리’이며, 전체 텍스트 문맥상 북극의 소수민족 문제는 러시아의 북극개발 과정에 있어 아주 작은 부분에 지나지 않다는 점을 보여주고 있다.

2020년에 발표한 ‘2035 러시아 북극 정책 기본원칙’에서 북극권 거주민의 삶의 질과 복지를 보장하고 의료지원과 인재 양성 및 북극 지역 거주민의 삶의 개선을 주요 과제로 삼고 있다. 또한 북극 지역 주민이 안정적으로 거주할 수 있는 사회 환경을 구축하는 것에 정책의 초점이 맞춰지고 있으나 이는 북극에 거주하는 일반 주민을 대상으로 하는 것으로, 북극개발에 의해 삶의 터전과 고유문화가 파괴되는 소수원주민들의 삶에 대해서는 고려가 없는 것으로 보인다. ‘타스 통신사’ 뉴스에 반영된 북극토착소수민족들에 대한 러시아 국민 및 정부의 태도를 감안하면, 이러한 정책시행으로 북극에 거주하는 토착소수민족들의 인구감소 및 민족문화의 해체나 소멸을 더욱 가속화시킬 것으로 보인다.

| 참고문헌 |

1. 논문 및 단행본

- 계용택 (2015). “북극에 관한 러시아 언론분석 및 한국의 대응전략.” 『한국 시베리아 연구』. 제19권. 2호, pp. 35-72.
- \_\_\_\_\_ (2020). “러시아 언론의 북극에 관한 보도내용 및 성향 분석: 타스통신사의 뉴스기사 텍스트를 중심으로.” 『한국 시베리아연구』. 제24권. 4호, pp. 31-59.
- 김일환 (2019). “인문학을 위한 신문 빅 데이터와 텍스트 마이닝.” 『어문론집』. 제78권, pp. 41-62.
- 김태종·박상욱 (2019). “뉴스 빅데이터를 활용한 평생교육 토픽 분석.” 『평생교육학 연구』. 제25권. 3호, pp. 29-63.
- 박진택 (2019). “R 텍스트 마이닝을 활용한 러시아 진보계열 미디어의 ‘크림반도 합병’ 관련 기사 분석: ‘RadioFreeEurope/RadioLiberty’, ‘Новая газета’의 2014-2015년, 2019년 기사 비교를 중심으로.” 『서울대학교 2019학년도 유럽지역학 연계진공 우수논문』, pp. 1-63.
- 박대민 (2013). “뉴스 기사의 빅데이터 분석 방법으로서 뉴스정보원 연결망 분석.” 『한국언론학보』. 제57권. 6호, pp. 234-262.
- \_\_\_\_\_ (2016). “장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성: 100만 건 기사의 정보원과 주제로 본 신문 26년.” 『한국언론학보』. 제60권. 5호, pp. 353-407.
- 박만희 (2016). “동시단어분석을 이용한 품질경영분야 지식구조 분석.” 『품질경영학회지』. 제44권. 2호, pp. 389-408.
- 유예림 (2017). “빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론 보도 분석.” 서울대학교 박사학위논문.
- 이은별·전진오·백지선 (2017). “서울의 다문화 공간 연구: 뉴스 빅데이터 분석 시스템 ‘빅카인즈’를 이용한 국내 언론의 외국인 마을 보도(1990-2016) 분석.” 『미디어 경제와 문화』. 제15권. 2호, pp. 7-43.
- 전채남·서일원 (2013). “빅데이터 분석의 기술마케팅 활용에 관한 연구: 잠재 수요기업 발굴을 중심으로.” 『마케팅논집』. 제21권. 2호, pp. 181-203.
- 한승희 (2019). “도서관에 대한 언론 보도 경향: 1990-2018 뉴스 빅데이터 분석.” 『정보관리학회지』. 제36권. 3호, pp. 7-36.
- Blei, D. M., A. Y. Ng and M. I. Jordan (2003). “Latent Dirichlet allocation” *Journal of machine Learning research*. Vol. 3. No. 1, pp. 993-1022.

- RADJABOVA, DILOROM (2011). “러시아 신문에 보도된 한반도 뉴스 특성에 대한 연구: 2002-2010년의 이즈베스티야에 보도된 한반도 관련 기사 내용분석.” 이화여자대학교 석사학위논문.
- Salton, G. (1989). *Automatic text processing*. MA: Addison-Wesley.
- Salton, G. and M. J. McGill (1983). *Introduction to modern information retrieval*. New York, NY: McGraw-Hill.
- Wang, W. and R. Rada. (1998). “Structured hypertext with Domain Semantics.” *ACM Train Inform System*. Vol. 16, pp. 372-412.
- Wasserman, S. and Faust. K. (1994). *Social network analysis: method and applications*. New York: Cambridge University Press.

## 2. 기타

- 경제·인문사회연구회 (2021). 『러시아 북극개발전략과 연계한 북극진출 방안 연구』. 국무총리(국무조정실).
- 에너지경제연구원 (2016). “러시아의 북극지역 자원개발 동향과 전망.” 『세계 에너지 시장 인사이트』. 제16-7호, pp. 17-28.
- 러시아 통신사 타스. <https://tass.ru/>. (2022년 1월 20일 검색)

| 논문투고일 : 2022년 05월 02일 |

| 논문심사일 : 2022년 05월 25일 |

| 게재확정일 : 2022년 06월 08일 |

| ABSTRACT |

**Analysis of Russian media's Reports and  
Tendencies on Arctic minorities:  
Focusing on the Text of News Articles of  
'TASS News Agency'**

**Ke Yong Tek**

(Russia Research Center)

This paper analyzed the news published in Russia and tried to examine the thoughts of the Russian people and the government on minorities living in the Arctic. The results of the analysis help identify the actual intentions of the Russian people and the government toward minorities, and may also serve as basic data for predicting the possibility of extinction of indigenous minorities living in the Arctic.

In order to analyze the content of news articles, this study visualized the network structure of keywords and the contents of the entire news article text using a simultaneous word frequency analysis technique, and analyzed the content and structure of the entire news article text based on this. In this study, R's TM package was used for text mining.

The essential intention of the article was inferred by analyzing the text of the 'TASS News Agency' searched as 'Arctic and Ethnic' by word frequency (word cloud presentation) analysis, network analysis between words, correlation analysis, and n-gram analysis method. As a result of the analysis, it can be seen that the correlation between the words of the Arctic native minority is very low compared to the Arctic development project. These studies show that the key keywords in the entire article are 'Socioeconomic Development in Russia' and 'Language Rights of

Minorities in the Arctic', and that the issue of minorities in the Arctic is only a small part of Russia's Arctic development process in full text. Russia's "2035 Russian Arctic Policy Basic Principles" guarantee the quality and welfare of Arctic residents, and medical support, talent training, and improvement of Arctic residents' lives, but there seems to be no consideration for the lives of minority residents whose native culture is destroyed by Arctic development. Considering the attitude of the Russian people and the government toward the Arctic indigenous minorities, the implementation of this policy is expected to further accelerate the demographic decline and the dissolution or destruction of the ethnic culture of the Arctic.

- Key words: Russia, Arctic Development, Ethnic Minorities, Text Mining, News Analysis