

A Keyword Network Analysis on Health Disparity in Korea: Focusing on News and its application to Physical Education

Woo-Kyung Kim*

Abstract

This study aimed to analyze the keyword related to Health Disparity in Korea through the method of keyword network analysis and to establish a basic database for suggesting ideas for prospective studies in physical education. To achieve the goal, this study crawled co-occurred keyword with 'health' and 'disparity' from news casted in 20 different channels. The duration of the news was 3 months, from September 11th, 2018 to December 11th. The results are as follows. First, among the news during recent 3 months, there were 1,383 keyword related to health disparity and this study selected 173 keyword which had co-occurred over 3 times. Second, the inclusiveness of the network was 97.674% and the density was .038. Third, analyzing news related to health disparity, 'mortality' was the most co-occurred keyword and 'disparity', 'reinforcement', 'the most', 'health', '6 times', 'Seoul', 'half', 'medicine', and 'local' were shown similarly. And common keyword in 4 centrality were 13 keyword. Lastly, by analyzing eigenvector centrality, significantly different result has shown. 'Disparity' was the most co-occurred keyword. Based on this result, this study showed the necessity for reinforcing the public physical education in public education system in Korea. In order to achieve it, the field of physical education must look beyond present elite-focused physical education to public physical activity.

▶ Keyword: Health, Disparity, Keyword Analysis, News, Physical Education

I. Introduction

인간이 건강하게 자신의 삶을 영위한다는 것은 매우 중요한 일이지만, 동시에 지키기 어려운 것 중 하나이기도 하다. 기초적인 위생 규칙을 지키지 못해 많은 사람들이 죽어가는 상황을 극복하기 위해, 전 세계가 하나가 되어 인간이 건강하게 살 수 있는 세상을 구축하기 위한 노력을 위해 WHO가 설립된 것이 1948년이므로, 인간 건강의 유지가 정부 그리고 세계적인 통제를 받을 수 있게 된 것 역시 그리 오래된 일이라고 하기는 어렵다. 이러한 인식을 반영하듯 Sen과 같은 보건학자들은 인간의 건강을 “인간 삶의 가장 중요한 조건 가운데 하나이며 인간능력의 핵심적인 구성요소”[1]로 정의하고 인간의 건강이 인권의 차원에서 가장 기초적이고 필수적인 요소라는 점을 강조한다. 이렇듯 건강은 인간의 가장 기초적인 상태를 지향하는 욕구이며, 기본적인 권리라고

할 수 있다. 인간이 소유하고자 하는 물질적인 것에 대한 욕심과는 달리, ‘건강이 좋지 않다’는 표현에 담긴 의미는 그 사람의 생명을 유지하기 위한 조건과 관련되어 있기 때문에 적극적으로 보호받아야 하는 권리라고 할 수 있다[2].

그러나 현실적으로 개인의 건강이 동등한 수준으로 유지되는 것은 아니다. 오히려 개인의 건강 상태는 인구의 숫자만큼이나 다양하다. 이러한 차이는 성별에 따른 차이일 수도 있고, 연령이 증가함에 따른 자연스러운 차이일 수도 있다. 하지만 사회 경제적 지위 역시 건강에 영향을 미치는 한 요인이 될 수 있다 [3]. 현대의 국가들은 대부분 자본주의 경제체제를 도입하고 있기 때문에, 경제체제 내에서 구성원 간의 불가피한 격차는 있을 수밖에 없고[4, 5], 이러한 경제적 격차가 건강상의 격차에

• First Author: Woo-Kyung Kim, Corresponding Author: Woo-Kyung Kim
*Woo-Kyung Kim (wkkim@anyang.ac.kr), College of Liberal Art, AnYang University
• Received: 2019. 01. 07, Revised: 2019. 02. 15, Accepted: 2019. 03. 04.

까지 이르러 사회적인 문제가 될 수 있다는 점이 몇몇 학자들
에 의해 연구된 바 있다[6, 7]. 따라서 현대 사회에서 건강 격
차를 줄이는 문제는 단순히 한 국가 내의 문제라기보다는 전
세계적인 문제라고 할 수 있다.

물론 과학기술의 발전과 더불어 현대인의 건강수준은 비약
적으로 향상되었다고 할 수 있다[8]. 그러나 여전히 잔존하는
경제적 격차 등으로 인해 실질적인 건강격차는 계속해서 보고
되고 있고[9], 이는 대한민국의 경우도 마찬가지이다. 최근의
지역 건강불평등 현황을 분석한 연구에 따르면 기대수명과 건
강수명, 소득 5분위 간 기대수명 및 건강수명 격차가 전국 252
개 시·군·구별로 큰 차이를 보이는 것으로 나타났다[10]. 이 연
구는 통계청 자료를 활용하여 지역별 연령표준화 사망률을 비
교하고 있는데, 수도권과 그 외의 지역의 사망률 격차가 여전히
줄어들지 않고 있고, 사망수준이 높은 지역과 낮은 지역이 뚜렷
하게 구분되는 현상이 나타났다는 연구와도 같은 맥락의 결과
를 보여주고 있다[11]. 이러한 연구들을 토대로 판단할 때, 그
동안 우리 사회가 많은 발전을 거듭해 왔음에도, 여전히 사람들
의 유전적인 개인차, 혹은 사회경제적 격차로 인한 건강 격차를
해결하지 못하고 있다는 사실을 알려주고 있다. 결국 이러한 건
강 격차는 사회, 경제, 제도 등의 구조적 요소들로 인해 개인,
집단, 지역 간의 건강에 차이가 발생하는 것을 말한다[12]. 따
라서 이러한 차이를 줄이기 위한 다양한 방면의 연구 및 정책
개발이 요구된다.

하지만 많은 논문들이 보건 및 사회복지 정책차원에서만 이
주제를 다루고 있을 뿐, 건강격차 문제는 지금까지 체육계에서
관심을 받지 못하였다. 오로지 신자유주의와 학교체육정책의
연관성을 논한 연구[13] 에서 신자유주의가 가지고 있는 몇몇
특성이 체육수업에 영향을 미쳐 건강격차를 발생시킬 수 있다
는 결과를 찾을 수 있을 뿐이었다. 그럼에도 해당 연구는 체육
학이 건강격차를 줄이는 데 중요한 역할을 수행할 수 있다는
단서를 제공해 준다. 건강격차 자체는 유전적인 요인 및 사회경
제적 요인을 통해 발생할 수 있으나, 정규 교육과정 내의 체육
수업을 통해 이러한 건강격차를 줄이는 데 공헌할 수 있다는
점이다. 결국 이러한 논리는 현재의 엘리트 체육 중심으로 투자
하고 있는 체육정책을 넘어, 생활체육을 대중화하는 것이 국민
의 건강격차를 줄이는 데도 기여할 수 있다는 점이다.

앞서 언급한 바와 같이 건강 격차를 다루는 체육학 연구가
거의 없는 실정에서, 사람들이 건강 격차에 대해서 어떤 의견을
가지고 있는지, 건강 격차는 어떤 키워드들과 함께 등장하는지
를 분석하는 것은 체육학 연구자들에게 새로운 연구 방향을 제
시해 줄 수 있을 것이다. 이에 본 연구에서는 국내 뉴스 속 건
강 격차가 어떤 관련 속에서 논의되고 있으며 어떤 키워드들이
중심성을 가지고 있는지에 대한 경향 분석을 통해 현대인들이
건강 격차를 해소하기 위한 기초자료를 제공하고자 한다.

II. Materials and Method

1. Research Procedure

본 연구는 건강격차에 대한 국내 뉴스를 분석함으로써 체육
계가 건강격차 해소에 기여할 수 있는 연구를 수행하기 위한
기초자료를 구축하고자 하는 시도이다. 건강격차를 다루는 뉴
스들의 동향을 파악하기 위하여 키워드를 중심으로 구조화된
형태의 정보를 추출하여 건강격차와 관련된 주제들의 패턴과
의미체계를 드러내고자 하였다. 이를 위해 다음과 같은 연구문
제를 설정하였다.

첫째, 건강격차와 관련된 뉴스기사들은 어떠한 키워드의 노
출빈도가 높은가?

둘째, 추출된 키워드들 간에 어떠한 관계가 있는가?

상기의 목적을 달성하기 위해서 국내에서 많이 활용되는 텍
스톰을 활용하여, 20개 뉴스채널을 대상으로 텍스트를 수집하
고, 키워드를 정제 및 분석/시각화하여 연구를 수행하였다.

첫째, 20개의 뉴스채널을 활용하여 키워드를 수집(crawling)
하였고, 검색 대상은 내용전체(타이틀+ 바디)가 아닌 타이틀을
기준으로 하였다.

둘째, 수집된 데이터는 비정형데이터이기 때문에 이를 분석
에 적합한 데이터로 변환하기 위한 정제 과정을 수행하였다.

셋째, 정제 과정을 거친 키워드들을 matrix 방식으로 조합하
여 원모드 데이터(one-mode data)로 변환하였다.

넷째, 키워드들의 빈도분석 및 네트워크 분석을 실시하였다.
네트워크 분석은 Netminer 4.0를 활용하여 키워드 간에 발생
하는 중심성을 다양한 각도로 분석하였다.

2. Materials and samples

본 연구는 키워드 네트워크 분석을 적용하여 국내 뉴스에서
건강격차가 어떤 주제들과 관련하여 논의되고 있으며, 어떠한
키워드들의 중심성을 가지고 있는지에 대한 언론의 경향을 탐
색하려는 목적을 갖는다. 이를 위해 본 연구는 연구의 대상과
자료수집 방식은 다음과 같이 설정하였다..

첫째, 텍스트에서 제공하고 있는 뉴스채널을 대상으로 키워
드를 추출하였다. 해당 프로그램에서 제공하고 있는 뉴스채널
은 KBS, MBC, SBS, YTN, 조선일보, 중앙일보, 동아일보, 한
겨레, 경향신문, 한국일보, 서울신문, 연합뉴스, NEWS1,
NEWSIS, oh my News, 노컷뉴스, 매일경제, 한국경제, 전자신
문, ZDNet Korea 총 20개 채널이다.

둘째, 상기의 20개의 뉴스채널에서 키워드를 크롤링
(crawling)하였다. 키워드는 '건강격차' 키워드와 공출현하는
키워드를 중심으로 하였고, 검색시점은 텍스트에서 크롤링을
지원하는 최대 기간 즉, 3개월(2018년 9월 11일부터 12월 11
일까지)로 설정하였다.

셋째, 연구에 적합하지 않은 키워드를 삭제하고, 유사 키워
드를 통합하는 과정을 거쳤다. 명사를 제외한 조사 등은 삭제하
였고, 의미가 중첩되거나 유사한 키워드를 통합하였다.

넷째, 이렇게 정제한 키워드는 원 모드 데이터(one-mode data) 중 매트릭스(matrix)로 변환하였다.

3. Method

건강격차와 관련된 키워드 간 네트워크를 분석하기 위하여 NetMiner 4.0을 활용하였다. 키워드 네트워크 분석은 키워드 사이의 관계를 분석하는 데 있어, 전체 네트워크 전체의 특성 및 개별 키워드들의 특성을 객관적으로 보여주는 분석방법이다[9]. 키워드 네트워크 분석의 항목은 결속(연결정도: degree, 밀도: density, 포괄성: inclusiveness, 연결강도: strength, 연결지속기간: duration), 중심성(centrality), 등위성(equivalence), 구조적 공백(structural hole) 등 다양하다[10]. 본 연구에서는 키워드의 밀도와 중심성(연결, 근접, 매개, 위세)을 분석하였다.

밀도(density) 항목은 네트워크 내에서 키워드들이 연결된 정도, 다시 말해 네트워크에 포함된 키워드들이 모인 정도를 의미한다[14]. 네트워크 내에 속한 키워드들이 상호 관계가 많을 때 밀도가 높다고 볼 수 있다. 네트워크 내의 키워드 밀도가 높으면 정보의 흐름이 빈번하게 발생한다는 것을 의미한다. 따라서 네트워크의 구조적 특성에서 밀도가 보여주는 것은 네트워크가 얼마나 완벽하게 구축되어 있는지의 여부이다. 이러한 개념은 네트워크 소시오그램 상에서 점들 간 라인(line)의 많고 적음으로 나타난다. 밀도는 네트워크 안에서 키워드들 사이의 가능한 모든 관계들 중, 실제로 연결되는 관계의 비율을 말하고, 0과 1사이의 값으로 이를 표현한다. 0은 아무런 관계가 없다는 것이고, 1은 모든 키워드들과 관계를 갖고 있음을 나타낸다.

중심성(centrality)은 한 키워드가 전체 네트워크에서 중심에 위치하는 정도를 나타내는 지표이다. 이를 통해 어떤 노드(Node), 즉 어떤 키워드가 네트워크 내에서 가장 중요한 키워드인지 찾아내고, 해당 네트워크 구조가 어떤 노드에 집중되어 있는지의 정도를 파악할 수 있는 개념이다[15]. 이러한 중심성은 연결정도중심성, 근접중심성, 매개중심성, 위세중심성으로 구분된다. 먼저 연결정도중심성(degree centrality)은 다른 키워드와 연결된 정도가 네트워크 내 중심에 얼마나 가까운지 파악하는 개념이다. 이는 한 키워드가 중심에 위치하는 정도를 계량화하는 것이다. 즉, 연결된 키워드가 많고 적음의 여부가 기준이 된다. 근접중심성(closeness centrality)은 한 키워드가 다른 키워드에 어느 정도 가깝게 있는지를 말하는 개념이다. 근접중심성이 높다는 것은 네트워크 내에서 다른 키워드들보다 빠른 경로를 통하여 전체 네트워크 내에 내재된 정보를 확보하거나 정보에 접근하기 쉽다는 것이다. 매개중심성(betweenness centrality)은 하나의 키워드가 최단거리 내에서 다른 키워드들을 매개하는 역할을 하고 있는 정도를 말한다[16]. 이는 한 키워드가 네트워크 내에서 매개자역할을 어느 정도 수행하는지를 측정하는 개념이다. 이 키워드는 전체 네트워크의 정보흐름에 있어서 중요한 매개역할을 하기 때문에 매개중심성을 측정하는 것은 네트워크의 구조적 특성을 파악하는 데 매우 유리하다. 하나의 키워드가 다른 키워드들과 연계되는 빈도값이 1에 가까울수록 중심성이 높은 키워드이며, 이 키워드는 전체 네트워크에서 그만큼

영향력을 많이 행사하게 된다. 마지막으로 아이겐벡터 중심성(Eigenvector centrality)은 하나의 노드에서 키워드가 다른 키워드들과 많이 연결되어 있을수록 높아지는 중심성을 말한다. 이것은 아이겐벡터 중심성이 높은 노드로부터 어떤 변화가 발생할 때, 그곳에서 발생하는 파급효과가 크다는 것을 의미한다.[17]

본 연구에서는 키워드 간의 동시출현(co-occurrence) 빈도를 근거로 삼아 키워드×키워드(1 mode) 매트릭스 데이터를 구축하여 네트워크 분석을 실시하였다.

III. Results

1. Result of the keyword analysis

20개의 뉴스채널에서 2018년 9월 11일부터 12월 11일까지 검색된 건강격차 키워드를 분석하였다. 수집 데이터인 비정형데이터를 분석용 데이터로 활용하기 위하여 텍스트 정제 과정을 수행했다. 이 과정에서 명사를 제외한 조사 등은 삭제하였고, 유사단어의 경우 데이터 정제를 통해 정리하였다. 이러한 데이터 정제를 통해 1,383개의 키워드를 도출하였다. 1,383개 키워드 중에서 10회 이상 공출현 하는 키워드는 7개(건강, 사망률, 격차, 한국, 삶, 개최, 6배)이었고, 9회 출현 키워드는 3개(불평등, 수준, 내년), 8회 출현은 5개(강화, 환자, 종합, 1위, 비정규직), 3회 이상 7회 이하 키워드는 164개로 확인되었다. 본 연구에서는 3회 이상 공출현 하는 키워드 173개를 분석 키워드로 설정하였다.

가장 높은 공출현 빈도를 보이는 키워드는 '건강'(14회)으로 분석되었고, '사망률', '격차', '한국'이 13회로 두 번째로 높은 빈도를 보였다. 세 번째로 높은 빈도를 보인 키워드는 '삶'(11회)이었고, '개최', '6배'(10회)의 순으로 분석되었다. 다음은 6회 이상 출현한 키워드를 그림으로 나타낸 것이다<Fig. 1>.

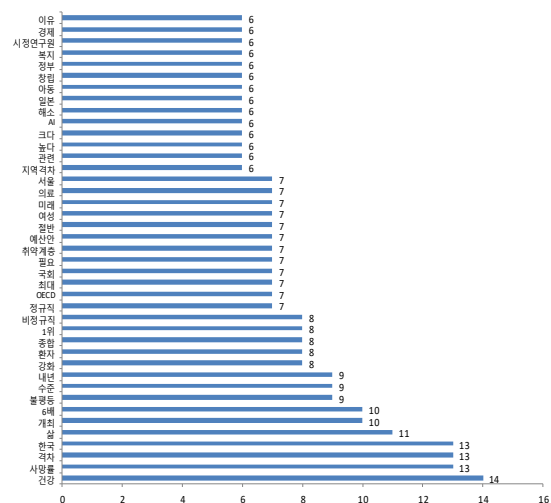


Fig. 1. Frequency of the co-occurrence

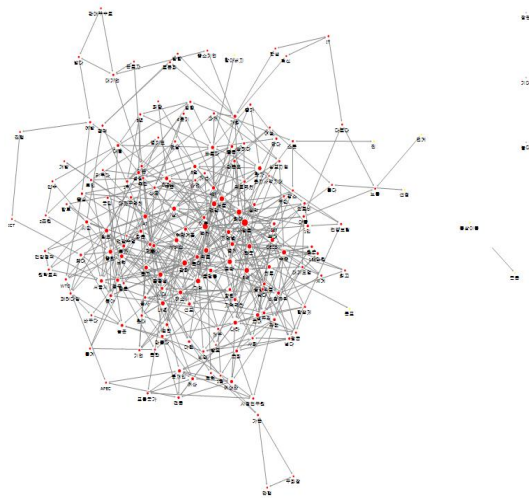


Fig. 2. Whole Keyword Network

2. Result of the network analysis

<Fig. 2>에서 제시된 바와 같이 본 분석은 건강 격차가 공출현하는 뉴스 기사에서 키워드를 추출하였고, 네트워크 분석결과 포괄성 (inclusiveness)이 97.674%로 나타났다. 본 연구에서 드러난 네트워크는 방향성이 없는 키워드들의 관계로써, 고립(isolate) 키워드는 4개, 전달(transmitter) 및 수신(receiver) 키워드는 존재하지 않았다. 그리고 매개(carrier) 키워드는 9개, 정상(ordinary) 키워드가 159개로 분석되었다. 키워드들의 전체 네트워크 밀도 (density)는 .038로 분석되었다.

다음은 연결중심성을 분석한 결과이다. 키워드 네트워크의 연결중심성 지수는 11.443%로 분석되었고, 연결중심성이 높은 상위 30개에 해당하는 키워드는 다음과 같다.

‘사망률’ 키워드의 중심성이 가장 높고, ‘격차’, ‘강화’, ‘최대’, ‘건강’, ‘6배’, ‘서울’, ‘절반’, ‘의료’, ‘지역’, ‘한국’, ‘수준’, ‘공공성’, ‘삶’, ‘내년’, ‘서비스’, ‘종합’, ‘AI’, ‘환자’, ‘OECD’, ‘필요’, ‘해소’, ‘복지’, ‘경제’, ‘나라’, ‘지역별’, ‘불평등’, ‘미래’, ‘높다’, ‘노력’의 순으로 분석되었다<Table 1, Fig. 3>.

Table 1. Degree Centrality of Keyword

Nr.	Keyword	Degree Centrality
1	death rate	0.1520
2	disparity	0.1287
3	reinforce	0.1111
4	maximum	0.1111
5	health	0.0994
6	6 times	0.0994
7	Seoul	0.0994
8	half	0.0936
9	medicine	0.0936
10	local	0.0936
11	Korea	0.0877
12	level	0.0877
13	publicity	0.0877
14	life	0.0819
15	next year	0.0819
16	service	0.0819

17	total	0.0760
18	AI	0.0760
19	patient	0.0702
20	OECD	0.0702
21	need	0.0702
22	solution	0.0702
23	welfare	0.0702
24	economy	0.0702
25	country	0.0702
26	regional	0.0702
27	inequality	0.0643
28	future	0.0643
29	high	0.0643
30	effort	0.0643

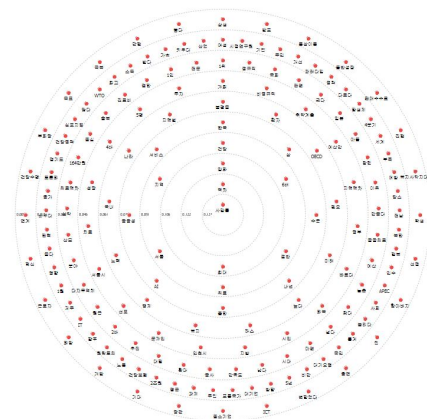


Fig. 3. Degree Centrality of Keyword

다음은 매개중심성을 분석한 결과이다. 키워드 네트워크의 매개중심성 지수는 10.611%로 분석되었고, 매개중심성이 높은 상위 30개에 해당하는 키워드는 다음과 같다.

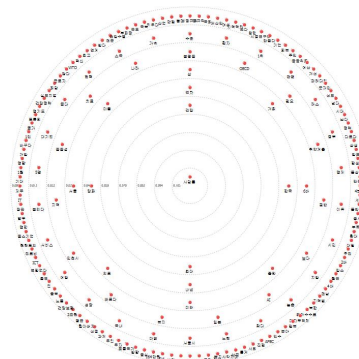


Fig. 4. Betweenness Centrality of Keyword

‘사망률’ 키워드의 중심성이 가장 높고, ‘최대’, ‘건강’, ‘강화’, ‘내년’, ‘격차’, ‘한국’, ‘개최’, ‘미래’, ‘삶’, ‘의료’, ‘종합’, ‘서울’, ‘아동’, ‘6배’, ‘빠르다’, ‘절반’, ‘일본’, ‘AI’, ‘지역’, ‘취약계층’, ‘인천시’, ‘복지’, ‘나라’, ‘필요’, ‘OECD’, ‘공공성’, ‘불평등’, ‘높다’, ‘치료’의 순으로 분석되었다<Table 2, Fig. 4>.

Table 2. Betweenness Centrality of Keyword

Nr.	Keyword	Betweenness Centrality
1	death rate	0.1170
2	maximum	0.0671
3	health	0.0634
4	reinforce	0.0583
5	next year	0.0569
6	disparity	0.0540
7	Korea	0.0519
8	hold	0.0462
9	future	0.0459
10	life	0.0449
11	medicine	0.0441
12	total	0.0397
13	Seoul	0.0393
14	movement	0.0376
15	6 times	0.0356
16	fast	0.0333
17	half	0.0326
18	Japan	0.0323
19	AI	0.0318
20	local	0.0298
21	poor class	0.0283
22	Incheon	0.0282
23	welfare	0.0274
24	country	0.0268
25	need	0.0257
26	OECD	0.0252
27	publicity	0.0248
28	inequality	0.0243
29	high	0.0243
30	treatment	0.0239

Table 3. Closeness Centrality of Keyword

Nr.	Keyword	Closeness Centrality
1	death rate	0.4193
2	disparity	0.4082
3	reinforce	0.4082
4	maximum	0.3866
5	health	0.3818
6	medicine	0.3799
7	Seoul	0.3762
8	half	0.3752
9	publicity	0.3734
10	service	0.3734
11	6 times	0.3725
12	life	0.3707
13	Korea	0.3680
14	regional	0.3645
15	inequality	0.3628
16	local	0.3628
17	AI	0.3611
18	next year	0.3594
19	future	0.3586
20	OECD	0.3577
21	high	0.3520
22	treatment	0.3496
23	5 people	0.3488
24	welfare	0.3480
25	regional disparity	0.3464
26	economy	0.3464
27	2 times	0.3457
28	total	0.3441
29	solution	0.3441
30	investment	0.3441

다음은 근접중심성을 분석한 결과이다. 키워드 네트워크의 근접중심성 지수는 24.188%로 분석되었고, 근접중심성이 높은 상위 30개에 해당하는 키워드는 다음과 같다.

‘사망률’ 키워드의 중심성이 가장 높고, ‘격차’, ‘강화’, ‘최대’, ‘건강’, ‘의료’, ‘서울’, ‘절반’, ‘공공성’, ‘서비스’, ‘6배’, ‘삶’, ‘한국’, ‘지역별’, ‘불평등’, ‘지역’, ‘AI’, ‘내년’, ‘미래’, ‘OECD’, ‘높다’, ‘치료’, ‘5명’, ‘복지’, ‘지역격차’, ‘경제’, ‘2배’, ‘종합’, ‘해소’, ‘투자’의 순으로 분석되었다<Table 3, Fig. 5>.

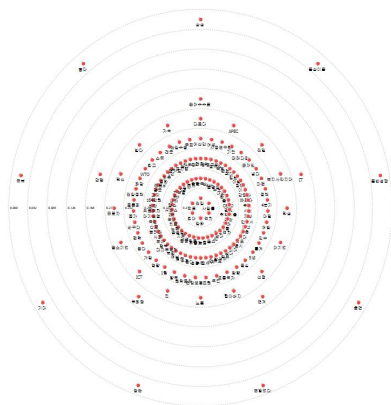


Fig. 5. Closeness Centrality of Keyword

마지막으로 아이겐벡터중심성을 분석한 결과이다. 키워드 네트워크의 아이겐벡터중심성이 높은 상위 30개에 해당하는 키워드는 다음과 같다.

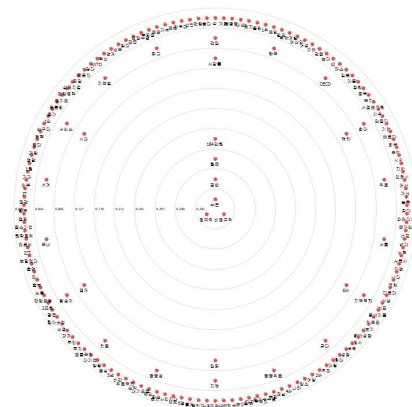


Fig. 6. Eigenvector Centrality of Keyword

‘격차’ 키워드의 중심성이 가장 높고, ‘공공의료’, ‘수준’, ‘절반’, ‘OECD’, ‘사망률’, ‘경제’, ‘크다’, ‘국내’, ‘강화’, ‘6배’, ‘시대’, ‘월급’, ‘세계’, ‘지역’, ‘의료’, ‘서비스’, ‘공공성’, ‘한국’, ‘건강’, ‘해소’, ‘지역별’, ‘최대’, ‘서울’, ‘항생제’, ‘정규직’, ‘비정규직’, ‘164만원’, ‘치료’, ‘최고’의 순으로 분석되었다<Table 4, Fig. 6>.

Table 4. Eigenvector Centrality of Keyword

Nr.	Keyword	Eigenvector Centrality
1	disparity	0.4244
2	public medicine	0.4228
3	level	0.3831
4	half	0.3770
5	OECD	0.3311
6	death rate	0.2945
7	economy	0.1247
8	bigger	0.1128
9	domestic	0.0980
10	reinforce	0.0963
11	6 times	0.0960
12	era	0.0915
13	income	0.0809
14	world	0.0772
15	local	0.0716
16	medicine	0.0716
17	service	0.0687
18	publicity	0.0649
19	Korea	0.0628
20	health	0.0624
21	solution	0.0596
22	regional	0.0595
23	maximum	0.0568
24	Seoul	0.0539
25	anti-biotic	0.0531
26	full-time	0.0514
27	part-time	0.0502
28	1.64 million won	0.0497
29	treatment	0.0459
30	highest	0.0441

네 개의 중심성에 공통으로 존재하는 키워드는 13개(격차, 강화, 건강, 최대, 절반, 6배, 서울, 의료, 한국, OECD, 공공성, 지역, 사망률)이고, 세 개의 중심성에 공출현하는 키워드는 12개(내년, 경제, 삶, 서비스, 지역별, 복지, 불평등, 종합, 미래, 높다, AI, 치료)로 분석되었다. 그리고 두 개의 중심성에 공출현하는 키워드는 5개(수준, 지역격차, 나라, 필요, 해소)로 나타났고, 22개(환자, 노력, 비정규직, 월급, 5명, 2배, 164만원, 투자, 시대, 세계, 국내, 크다, 항생제, 공공의료, 최고, 개최, 아동, 빠르다, 일본, 취약계층, 인천시)의 키워드는 상위 30개의 중심성에 1번 출현하는 것으로 분석되었다. 정리하면 네 개의 중심성에 공통적으로 상위 30위에 들어가는 키워드는 총 52개로 나타났다.

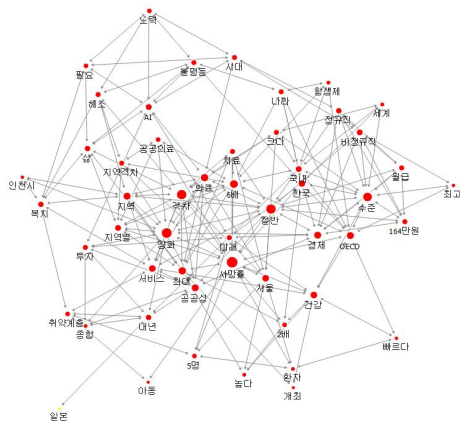


Fig. 7. Keyword Network of Top 30

네 가지 중심성에서 상위 30개에 속하는 키워드들의 합집합인 52개 키워드들의 네트워크는 <Fig. 7>과 같다. 네트워크 분석결과 포괄성(inclusiveness)은 100%이었고, 고립(isolate), 전달(transmitter), 수신(receiver) 키워드는 없었고, 매개(carrier) 키워드 1개, 정상(ordinary) 키워드가 51개로 분석되었다. 키워드들의 전체 네트워크 밀도(density)는 .142로 분석되었다.

IV. Discussion

본 연구는 키워드 네트워크 분석을 적용하여 국내 뉴스에서 건강격차가 어떤 주제들과 관련하여 논의되고 있으며, 어떠한 키워드들의 중심성을 가지고 있는지에 대한 언론의 경향을 탐색하려는 목적을 갖는다. 이를 위해 본 연구는 총 20개 채널을 대상으로 키워드를 수집하였다. 키워드는 ‘건강격차’ 키워드와 공출현하는 키워드를 중심으로 하였고, 검색시점은 텍스트에서 크롤링을 지원하는 최대 기간 즉, 3개월(2018년 9월 11일부터 12월 11일까지)로 설정하였다. 이후 직접 검토를 통해 연구에 적합하지 않은 키워드를 삭제하고, 유사 키워드를 통합하는 과정을 거쳤다. 이렇게 구축된 자료를 이용해 네트워크 분석을 실시하였다.

앞서 밝힌 바와 같이 건강격차 자체는 유전적인 요인 및 사회경제적 요인을 통해 발생할 수 있으나, 정규 교육과정 내의 체육수업을 통해 이러한 건강격차를 줄이는 데 공헌할 수 있다. 이러한 논리에 기반하여 본 연구는 현재의 엘리트 체육 중심으로 투자하고 있는 체육정책을 더욱 발전시켜 생활체육을 대중화할 필요를 주장하고자 하는 목적을 갖는다. 이러한 점에서 과연 사람들이 건강 격차에 대해서 어떤 의견을 가지고 있는지, 그리고 건강 격차는 어떤 키워드들과 함께 등장하는지를 분석하는 것은 체육학 연구자들이 앞으로의 연구 방향을 모색하는 데 도움을 줄 수 있을 것이기 때문이다. 그 분석 결과는 다음과 같다.

첫째, 20개의 뉴스채널에서 2018년 9월 11일부터 12월 11일까지 검색한 건강격차 키워드를 분석한 결과 총 1,383개의 키워드를 도출하였다. 본 연구는 1,383개 키워드 중에서 3회 이상 공출현하는 키워드 173개를 분석 키워드로 설정하였다.

173개의 키워드 중 가장 높은 공출현 빈도를 보이는 키워드는 ‘건강’(14회)으로 분석되었고, ‘사망률’, ‘격차’, ‘한국’이 13회로 두 번째로 높은 빈도를 보였다. 세 번째로 높은 빈도를 보이는 키워드는 ‘삶’(11회)이었고, ‘개최’, ‘6배’(10회)의 순으로 분석되었다. ‘건강’, ‘격차’, 그리고 ‘삶’이라는 키워드가 높은 빈도로 공출현 한다는 점은 특이한 결과는 아니다. 그럼에도 ‘사망률’이라는 키워드가 두 번째 수준으로 높은 빈도로 공출현 한다는 결과는 언론들이 건강 격차와 관련하여 사망률을 가장 중요한 주제로 생각하고 있다는 것을 나타낸다. 이와 관련하여 ‘한국’이라는 키워드는 건강 격차와 사망률의 관계를 다른 국가와의 비교 속에서 논의하고 있다는 점을 보여준다. 또한 ‘개최’와 ‘6배’라는 키워드가 높은 빈도로 공출현 했다는 점은 의미심장하다. ‘개최’의 경우, 지난 9월에는 남북정

상회담 개최와 관련하여 남북교류에 있어서의 건강격차 해소가 이슈가 되었고, 11월에는 지역 간 건강격차의 원인규명 및 해소를 위해 질병관리본부와 지역보건 분야 학회들 간의 공동 협약을 개최하였다. 이와 관련하여 논의된 것이 치료 가능한 사망률의 지역 격차가 '3.6배'라는 점, 그리고 남북의 기대수명격차가 '7.6배'라는 점이 반영되어 '6배'라는 키워드가 높은 빈도로 공출현한 것으로 판단된다.

둘째, 네트워크 분석결과 포괄성(inclusiveness)이 97.674%로 나타났다. 키워드들의 전체 네트워크 밀도(density)는 .038로 분석되었다. 이러한 결과는 키워드 네트워크를 표현한 <Fig. 2>의 모습과 비교해볼 때 몇몇 키워드들, 즉 앞서 언급한 30개의 키워드들 간의 밀도는 매우 높으나, 그 외 다양한 키워드들이 해당 주제와 관련하여 사용되어, 비교적 넓은 주제범위를 포괄하고 있다고 해석할 수 있다. 결국 30개 키워드들 간의 정보 교환이 가장 활발하며, 이렇게 모인 키워드들 관계 바깥에서 밀도의 평균을 낮추는 키워드들은 각 언론사 별 특징 혹은 기자들 사이의 차이를 나타낸다고 할 수 있다. 그럼에도 포괄성은 매우 높게 나타나 있기 때문에 건강과 격차에 관련한 논의는 다양한 각도로 활발하게 보도되었다는 것을 의미한다.

셋째, 연결중심성, 매개중심성, 근접중심성의 상위 30개 키워드 순위를 분석한 결과, 공통적으로 '사망률' 키워드의 중심성이 가장 높게 나타났고, '격차', '강화', '최대', '건강', '6배', '서울', '절반', '의료', '지역' 등의 키워드는 큰 순위 차이 없이 유사하게 나타나는 것으로 분석되었다. 이는 순위가 높은 키워드들이 서로 공출현하는 빈도가 매우 높다는 것을 의미하는데, 건강 격차를 다룸에 있어서 가장 중요한 기준으로 삼는 지표가 '사망률'이라는 점, 그리고 서울과 지역 간의 사망률 차이가 확연한 점 등이 반영된 것으로 이해할 수 있다.

또한 공통으로 나타난 키워드들을 살펴보면, 총 네 가지 중심성에 공통으로 존재하는 키워드는 13개(격차, 강화, 건강, 최대, 절반, 6배, 서울, 의료, 한국, OECD, 공공성, 지역, 사망률)이고, 세 개의 중심성에 공출현하는 키워드는 12개(내년, 경제, 삶, 서비스, 지역별, 복지, 불평등, 종합, 미래, 높다, AI, 치료)로 분석되었다. 그리고 두 개의 중심성에 공출현하는 키워드는 5개(수준, 지역격차, 나라, 필요, 해소)로 나타났다. 이러한 결과는 건강 격차를 다룸에 있어서, 언론들이 '공공성'을 매우 중시하는 것으로 해석할 수 있다. 결국 건강 격차를 줄이는 사업은 어떤 유전적인 요인 등에 대한 대처라기 보다는, 국가 전체가 국민 전체를 대상으로 수행해야 하는 사업이라고 이해하고 있다는 것을 알 수 있다. 그리고 'OECD'와 '한국' 그리고 '서울'과 '지역'이라는 키워드가 공통적으로 등장하는 점으로 미루어 '한국'과 선진국들의 모임, 그리고 '서울'과 '지역'이라는 큰 범위에서 계속해서 건강 격차를 비교하고 있음을 보여준다. 결국 이러한 경향은 건강 격차가 더 이상 특정 지역만의 문제가 아니라 전국 그리고 전 세계의 문제라는 점을 보여주고 있으며, 건강이라는 개념이 기존에 생존과 관련된 절대적 관점의 문제에서, 상대적인 격차를 중요시하는 상황으로 변화했다고 해석할 수 있다.

넷째, 반면 아이겐벡터중심성을 분석한 결과는 앞의 세 가지

중심성을 분석한 결과와 매우 상이하게 도출되었다. 또한 키워드 순위 역시 앞의 세 가지 중심성에서의 순위와 다른 것으로 나타났다. 아이겐벡터 중심성은 연결된 다른 키워드들의 중요도까지도 반영한 결과이기 때문에, 결과가 상이하게 나타나는 것 자체가 문제는 아니다. 다만 아이겐벡터중심성을 분석한 결과 '격차' 키워드의 중심성이 가장 높고, '공공의료', '수준', '절반', 'OECD', '사망률', '경제', '크다', '국내', '강화', '6배', '시대', '월급', '세계', '지역', '의료', '서비스', '공공성', '한국', '건강', '해소', '지역별', '최대', '서울', '항생제', '정규직', '비정규직', '164만원', '치료', '최고'의 순으로 분석되었다. 이 순위는 앞서 논의한 세 가지 중심성들의 키워드 순위와 비교해 봤을 때, 결국 건강의 문제가 절대적인 생존의 문제라기보다는 삶의 질과 관련한 문제가 되었고, 따라서 상대적인 격차를 크게 중요하게 여기는 시대로 변화했다는 것을 보여준다. 이러한 결과는 결국 OECD와 대한민국, 그리고 서울과 지역 사이에 발생하는 건강 격차를 공공의료 수준을 높임으로써 해결하고자 한다는 결론이 가장 많이 보도되고 있다고 해석할 수 있다. 이러한 결과를 체육학 분야의 관점에서 해석한다면 결국 건강 격차를 해소하는 데 있어서 상대적인 격차를 중요하게 여기고, 이러한 것이 공공의료의 수준을 높임으로써 해결할 수 있는 것이라면, 앞서 인용한 연구[18]에서 드러난 바와 같이, 공공교육에서 체육활동의 비중을 높이고 생활체육의 활성화를 꾀하는 것이 건강 격차를 줄이는 데에도 도움이 될 수 있다는 논리적 귀결을 이끌어 낸다. 따라서 체육학 분야에서 공공교육에서 그리고 일상생활에서 사람들의 체육활동 참여를 더욱 증진할 방안을 지속적으로 탐구해야 할 필요성이 제기된다.

본 연구의 결론을 토대로, 다음과 같이 제언하고자 한다.

첫째, 본 연구는 '건강 격차'라는 키워드를 20개의 뉴스채널에서 2018년 9월 11일부터 12월 11일까지 보도한 뉴스들에서 검색하여 분석한 연구이다. 이러한 키워드 네트워크 분석은 빅데이터 분석 방식의 유행과 함께, 이미 세계적으로 다양한 분야에서 적용하여 수행하고 있는 방법론이며, 키워드 수집 기간을 달리하거나, 해당 키워드를 조작하는 방법에 따라 그 분석 내용과 해석은 모두 달라질 수 있는 여지를 가지고 있다. 이러한 점을 감안하여, 후속 연구에서는 전 세계 혹은 OECD국가들의 뉴스 및 연구들을 크롤링하여 건강 격차에 대해서 각국의 언론들은 어떤 입장을 가지고 있는지를 해석한다면, 향후 건강 격차를 줄이는 데 있어서 선도적인 연구를 수행하는 데 도움이 될 것으로 판단된다.

둘째, 본 연구는 키워드 네트워크 분석 방법론을 이용한 연구이다. 해당 방법론은 결국 빅데이터를 수집하여 얻어낸 자료를 프로그램 상에서 기계적으로 분석하는 방법이다. 하지만 해당 분석 결과에 대한 해석은 인간의 주관성을 거쳐 수행할 수밖에 없기 때문에, 실제로 건강 격차를 다루고 있는 뉴스이지만 건강 격차라는 키워드를 포함하지 않거나 다른 키워드들이 강세를 나타내는 경우에는 본 연구의 분석 대상에 포함되지 않았을 수 있다. 이는 인간의 언어, 그리고 사람들의 주목도를 높이기 위해 언론이 취하는 표현상의 다양성과 관련이 있고, 한국어 문법 상 주어가 생략되는 경우가 상당히 많기 때문이다. 본 연구의 분석 방법은 이러한 대상들이

배제될 가능성을 무시할 수 없다. 다만 본 연구에서는 분석 대상을 규정하는 데 있어 명확한 기준을 설정하기 위해 의도적으로 범주가 넓은 키워드를 선택하였다. 따라서 이러한 점을 보완하기 위해 후속 연구에서는 더욱 다양한 키워드 및 주제 선별 작업을 시도함으로써 주제에 더 적합한 분석 방식을 적용할 필요가 있다.

마지막으로, 본 연구는 ‘건강 격차’와 관련한 체육학 내부의 연구 성과가 제한된 상황에서, 보도된 뉴스들만으로 빅데이터를 활용한 네트워크 분석이다. 따라서 본 연구에서 분석한 내용들은 비교적 넓은 범위에서의 건강 격차 맥락을 담고 있을 뿐, 체육학과의 필연적인 접점까지 포함하고 있지는 않다. 따라서 후속 연구에서는 빅데이터 분석 방법을 다양한 플랫폼, 예를 들어 트위터 혹은 인스타그램 등에 해쉬태그로 표현된 내용까지 포함하여 분석하는 방식을 생각해볼 수 있다. 이러한 분석 방식에 약간의 개인 정보, 즉 성별과 연령대까지 고려하여 분석할 수 있다면, 본 연구의 결과에 더한 풍부한 해석이 가능할 것이다. 따라서 체육학 연구자들이 지속적인 분석 방법론을 탐구하고 데이터 수집 방법을 다변화함으로써 체육학계의 오랜 숙원인 생활체육 활성화를 이루기 위해 필요한 요건들을 도출할 수 있을 것이다.

REFERENCES

- [1] Sen, A., “Why Health Equity?,” *Health Economics*, Vol. 11, pp. 659–666, 2002.
- [2] J. G. Kim, “Regional Differences in the Equity of Health Care Utilization in Korea”, *J. Korean Social Security Association*, Vol. 28, No. 4, pp. 153–180, 2012.
- [3] H. S. Shin, D. J. Kim, “Health Inequality Measurement in Korea Using EuroQol-5 Dimension Valuation Weights”, *Journal of Preventive Medicine and Public Health*, Vol.41, No.3, pp. 165–172. 2008
- [4] N. Kakwani, A. Wagstaff and E. van Doorslaer, “Socioeconomic Inequalities in Health: Measurement, Computation, and Statistical Inference,” *Journal of Econometrics*, Vol. 77, pp. 87–103, 1997.
- [5] A. Wagstaff, P. Paci and E. van Doorslaer, “On the Measurement of Inequalities in Health”, *Social Science & Medicine*, Vol. 33, No. 5, pp. 545–557, 1991.
- [6] Y. H. Khang, J. W. Lynch, S. Happer, S. Yang, S., Jung-Choi, and H. R. Kim, “Material, psychosocial, and behavior factors to explain education and occupational mortality inequalities in a nationally representative sample of South Korea: relative and absolute perspectives”. *Soc Sci Med*, Vol. 68, pp. 858–866, 2009
- [7] P. Watts, D. Buck., G. Netuveli, and A. Renton, “Clustering of lifestyle risk behaviours among residents of forty deprived neighbourhoods in London: lessons for targeting public health interventions”. *Journal of Public Health*, Vol. 38, No. 2, pp. 308–315. 2015.
- [8] J. G. Kim, “The Changes in Income-related Health Inequality and Their Influencing Factors from 2005 to 2010”, *J. Korean Social Security Association*, Vol. 28, No. 3, pp. 121–151, 2012.
- [9] Y. Shin, T. H. Yoon, M. J. Kim, B. G. Jeong, B. G, and J. H. Seo, “Health equity policy and practice: Strategic approach for alleviation of health inequality.”, *Korea Social Policy Review*, Vol. 18, No. 4, pp. 41–77, 2011
- [10] J. W. Park, “Socioeconomic Inequalities in Health at the Regional Level in Korea’, *Health and welfare policy forum*, Vol. 260, pp. 7–19, 2018.
- [11] J. G. Kim, “Regional Differences in the Equity of Health Care Utilization in Korea”, *J. Korean Social Security Association*, Vol. 28, No. 4, pp. 153–180, 2012.
- [12] P. Braveman, “Health disparities and health equity: Concepts and measurement”. *Public Health*, Vol. 27, pp. 167–194, 2006.
- [13] H. W. Chung, Y. N. Park, and K. D. Lee, “Neo-Liberalism and Education Policy in School Physical Education: Finding its Effect and Educational Meaning”, *Journal of Korean Society of Sport Policy*, Vol 13, No. 4, pp. 1–18, 2015
- [14] D. W. Sohn, “*Social Network Analysis*”, Seoul: Kyungmun Sa. 2010.
- [15] S. Wasseman and K. Faust, “*Social Network Analysis: Methods and Applications*”, NY: Cambridge University Press. 1998.
- [16] D. W. Sohn, “*Social Network Analysis*”, Seoul: Kyungmun Sa. 2010.
- [17] D. W. Sohn, “*Social Network Analysis*”, Seoul: Kyungmun Sa. 2010.
- [18] H. W. Chung, Y. N. Park, and K. D. Lee, “Neo-Liberalism and Education Policy in School Physical Education: Finding its Effect and Educational Meaning”, *Journal of Korean Society of Sport Policy*, Vol 13, No. 4, pp. 1–18, 2015.

Authors



Woo-Kyung Kim received the doctoral degree of Physical Education from Sookmyung Women’s University. Dr. Kim joined the faculty of the College of Liberal Art at Anyang University, in 2012. She is currently an adjunct professor of the College of Liberal Art at Anyang University. She is interested in Measurement and Evaluation of Physical Education.