

교과 어휘의 의미 구조와 관계 변화 탐색

— 사회과 교과서 어휘의 워드 임베딩 및 네트워크 분석을
중심으로

전영주 이화여자대학교 에듀테크융합연구소 박사후연구원

- * 이 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF - 2022S1A5B5A16056943).

- I. 서론
- II. 연구 대상 및 방법
- III. 연구 결과 및 분석
- IV. 논의 및 시사점
- V. 결론

I. 서론

아동의 어휘 발달은 학령기 이후 학습을 통해 새로운 국면을 맞이한다. 구어적 상호작용을 통해 어휘를 습득하던 아동은 학교 교육을 통해 다양한 문자 언어 텍스트를 읽고 쓰며 일상에서 접할 수 없는 낯선 단어를 습득하며, 청소년을 거쳐 성인 초기에 이르면 약 6만 개의 단어를 이해하고 사용할 수 있게 된다고 한다(Nippold, 2006).

이때 어휘 발달은 단순한 양적 증가를 넘어, 언어가 추상적 개념과 복잡한 의미 체계를 구성하는 도구(Vygotsky, 1934/2011)로 발전하는 질적 변화를 포함한다. Nippold(2006), 배희숙(2021) 등에서는 학령기 아동 및 청소년 학습자의 어휘 발달이 중의적, 상징적, 은유적 의미 파악 능력이라는 질적 차원으로 나타날 수 있으며, 사회나 과학 등의 내용 교과 학습에서 경험하는 추상적 개념이나 전문 용어의 사용이 그 배경이 된다고 설명한다.

이처럼 교과 학습은 학습자 어휘의 양적·질적 발달의 계기가 되기도 하지만, 동시에 학습자의 어휘 수준은 교과 학습을 위한 문식성의 기반이 된다.¹⁾ 이에 국어과 교육에서는 어휘 교육을 체계적으로 논의하고자 ‘기초 어

휘’, ‘사고도구어’, ‘교과 개념어’ 등을 구분하고 목록화하였으며(김광해, 2003; 신명선, 2004; 전영주, 2019; 최소영, 2020 등), 이는 타 교과 학습에서 어휘의 중요성을 전제한 결과이기도 하다. 이러한 논의는 가르쳐야 할 어휘의 범주를 명확히 규정하고, 이독성 및 문해력 측정 등의 확장적 논의로(최소영·길호연·박태준·류수경, 2022) 이어지고 있다는 점에서 의의가 있다. 그러나 이를 교육적으로 실천하기 위한 어휘 능력과 발달 차원의 논의나 타 교과 학습과의 실제적 연계를 위한 국어과 어휘 교육 논의는 여전히 부족한 상태이다.

한편, 사회나 과학 등의 내용 교과에서도 어휘는 중요한 연구의 대상이다. 단어의 의미가 지식과 동일하지는 않지만, 지식이 단어로 표상된다는 점에서 교과 개념의 표상화 방식, 기술 방식 등에 주목하는 것이다. 특히 교과의 맥락에서 단어 간의 관계를 확인하는 것은 해당 교과의 지식 개념화 양상을 확인할 수 있는 근거가 된다. 사회과의 박세훈·장인실(2018), 이윤미·김효정·송원섭·조철기(2024), 과학과의 박서빈·전영석·이창훈(2015), 정동권·연유상(2021) 등은 개념이 단어의 축자적 의미에 국한되지 않고 교과 지식 체계에서 개념과 개념의 연결을 통해 구성된다는 관점에서 교과 어휘의 관계를 탐색한 연구이다. 이들 연구에 따르면, 학습자가 이전에 알던 단어의 의미는 불완전한 개념이다. 이에 교과의 지식 체계를 전제로 개념 변화를 유도하고 새로운 지식을 구성하는 교육적 방안을 고민하는 것이다.²⁾

-
- 1) 본 연구에서의 문식성은 교과 문식성(disciplinary literacy)을 말한다. 교과 문식성은 타 교과 및 다양한 주제 영역의 텍스트를 수용, 생산하는 데 필요한 언어적 관습성, 소통방식, 사고의 관습을 문식성의 전략적 도구로 삼으며 어휘를 문식성의 본질로 강조한다 (Shanahan & Shanahan, 2012). 근래 국어과 교육은 범교과 차원의 문식성의 한계를 논의하며 그 대안적 관점으로 교과 문식성에 주목한다. 관련 논의는 전영주(2020), 이용준(2021), 주세형(2021) 등을 참고할 수 있다.
 - 2) 이러한 접근은 구성주의 교육관에 기반한다. 예를 들어, ‘힘’이나 ‘일’, ‘에너지’와 같은 물리학의 개념을 일상의 개념으로 이해하거나 해석하는 것은 오개념(misconception)(박서빈 외, 2015)으로 간주되며 교정의 대상이 된다. ‘정치’의 개념을 일상에서 접하는 뉴스를 통해 이해하는 것은 선개념(preconception)이 된다(윤길복, 2012). 학습은 이러한 학습자의 기존 개념을 변화하게 하여 새로운 지식을 구성하는 과정이 되어야 한다는 것이다.

교과의 입장에서 지식 개념의 불완전한 이해는 교정이 필요한 대상이지만, 학습자 개인의 인지와 언어 발달의 관점에서 보면 이는 발달 과정의 일부이다. 지식을 표상하는 언어의 이해는 단순히 ‘단어’ 습득의 문제만이 아니라 언어와 문식성이 발달하며 지식의 구조와 맥락 속에서 그 의미를 점진적으로 구성하는 과정이기 때문이다(Fang, 2012). 이는 결과적으로 학습자 어휘의 질적 심화를 의미하며, 타 교과 학습을 위한 국어과 어휘 교육 역시 발달의 관점에서 ‘의미’를 중심으로 구체화될 필요가 있다.

교과의 지식 구성 과정이 학습자 어휘를 양적으로 확장할 뿐만 아니라³⁾ 의미를 발견하고 구성할 수 있는 맥락을 제공한다는 관점에서 본 연구는 사회과 교과 어휘의⁴⁾ 의미 구조와 관계 변화를 위드 임베딩과 네트워크 분석을 활용하여 탐색한다. 그 결과를 토대로 학습을 통한 어휘 발달을 의미 차원에서 재고하고, 국어과 교육에서 문식성에 기반이 되는 교과 어휘의 교육 방향을 제언하고자 한다.

II. 연구 대상 및 방법

1. 분석 대상

본 연구는 교과 어휘의 의미 구조와 관계 양상의 탐색을 위해 초등학교

-
- 3) Fitzgerald, Elmore, Relyea, & Stenner(2020)는 사회와 과학 교과서에 등장하는 영역 특수 어휘(domain specific vocabulary)의 최초 출현 시기가 학습자의 어휘 습득 연령과 유의미한 상관관계가 있음을 실증적으로 확인하였다. 여기서 영역 특수 어휘란 일반적인 사용 빈도는 낮지만, 특정 학문 영역의 지식 개념을 표상하는 단어를 말한다.
 - 4) 본 논의에서 말하는 ‘교과 어휘’는 교과의 지식 개념을 구성하는 과정에 관여하는 어휘를 포괄적으로 지칭한다.

3학년부터 고등학교 1학년까지의 사회 교과서의 어휘를 분석 대상으로 삼았다. 교과서는 교과 어휘로 구성된 텍스트로서, 학령기 아동 및 청소년이 일상에서 접하기 어려운 단어를 경험할 수 있는 자료가 된다. 또 학년이 올라가며 교육과정의 나선형 구조에 따라 반복·심화되는 어휘가 많으며, 그 의미 구조가 교과의 잠정적 지식 구조를 반영하면서도, 학습자의 연령과 인지적 수준을 고려한 비교적 객관적인 자료라 할 수 있다.

특히 사회과는 사회의 다양한 현상을 이해하고 관찰하기 위한 관점 을 제공하는 교과로서(윤길복, 2012), 그 교과 지식을 표상하는 어휘는 일상적 언어와의 관계를 통해 의미가 구체화되는 경우가 많다(전영주·김은성, 2015). 또한, 정치, 경제, 사회문화, 지리, 역사의 등 다양한 학문적 이론을 배경으로 각 영역이 독립적이면서도 밀접하게 관련된다는 점, 객관적 지식 개념의 이해보다는 해석과 판단을 강조한다는 점(윤길복, 2012)에서 교과 어휘의 의미 구조와 관계 변화를 확인하기 적절한 교과라 판단하였다.

본 연구는 분석 대상 어휘를 추출하기 위해 2015 개정 교육과정이 적용된 초중고 사회과 교과서 45권을 코퍼스화 하였다. 교과 학습이 독립적으로 시작되는 초등학교 3학년부터 공통교육과정인 고등학교 1학년까지의 사회과 교과서 중 비교적 인지도와 접유율이 높다고 알려진 세 출판사(비상, 미래엔, 천재)의 책을 원자료로 한 것이다. 초등학교 교과서에서 한국사를 별도로 다루는 단원과 학기(5학년 2학기)는 분석 대상에서 제외하였다.⁵⁾

교과서 텍스트는 한글 형태소 분석기인 Kiwi(이민철, 2024)를 활용하여⁶⁾ 토큰화하고 숫자나 기호, 영문은 정제 처리하였다. 각 코퍼스의 기본 계

-
- 5) 중학교와 고등학교 과정에서 한국사는 별도의 교과서로 제공된다. 초등학교 역사 영역을 함께 분석하면 중학교, 고등학교의 역사 및 한국사 교과서를 함께 고려하지 않을 수 없다. 논의의 범위가 지나치게 확장되지 않도록 역사 분야의 텍스트는 제외한 것이다.
 - 6) Kiwi는 사용자 사전에 공백을 포함할 수 있는 장점이 있다. 예를 들어 ‘온실 효과’와 같은 구 구성 용어는 공백을 없애는 전처리 작업 없이 사용자 사전을 사용하여 하나의 단위인 NNG로 태깅할 수 있다.

량 결과는〈표 1〉과 같다. 코퍼스는 교육과정의 학년군을 기준으로 초등학교 3, 4학년군(E34), 5, 6학년군(E56), 중학교(M), 고등학교(H) 4개로 구분한 것이며 총 120만 토큰의 코퍼스이다.⁷⁾ 워드 임베딩과 네트워크 분석에서는 교과 개념 지식의 의미화 양상을 확인하기 위해 분석 대상 단위를 일반명사(NNG)⁸⁾에 한정하였다.

〈표 1〉 사회과 교과서 어휘 코퍼스의 기본 계량 정보

학년군	총 빈도(Token)	총 유형(Type)	NNG 빈도	NNG 유형
E34	167,493	5,645	51,479	3,496
E56	231,205	7,416	76,391	4,751
M	486,645	11,854	178,025	8,007
H	354,088	10,974	124,139	7,806
Total	1,239,431	-	430,034	-

2. 워드 임베딩과 네트워크 분석

단어의 의미 구조와 관계 변화는 워드 임베딩과 네트워크 분석을 활용

-
- 7) 고등학교 코퍼스는 다른 학년군과는 달리 1년 과정의 교과서를 원자료로 하여 구성되었다. 두 개 학년의 교과서로 구성된 중학교보다 규모가 작을 수밖에 없다. 어휘의 양적 수준은 코퍼스 자체의 크기보다는 어휘 다양도(Type-Token Ratio) 등의 지표를 활용 한다. 참고로 각 코퍼스 어휘 다양도는 고등학교가 가장 높게 나타났다. Guiraud 지수(GI:Guiraud Index)(Daller, van Haut, & Treffers-Daller, 2003)를 활용하여 단원별로 어휘 다양도를 계산한 결과, 중학교 코퍼스는 10.19~18.04, 고등학교는 13.78~19.75로 계산되었다. 이는 고등학교가 중학교에 비해 저빈도의 단어가 다양하게 출현한 것으로 해석된다.
- 8) 여기서 일반명사는 토큰화되어 NNG 태그를 갖는 단위로, 용언의 명사 어근을 포함한다. 예를 들어, ‘증명하다’는 동사이나, 용언 어근을 분리하는 토큰화 방식을 선택하면 ‘증명/NNG + 하/XSN + 다/EF’로 결과가 출력된다. 〈표 1〉의 NNG 빈도는 용언의 명사 어근을 포함한 결과로, 코퍼스 원자료의 내용어를 대부분 포함한다.

하였다. 두 방법은 비정형 데이터인 텍스트에서 단어의 의미 관계나 개별 단어의 중심성을 수치화하며, 이를 기반으로 시각화, 의미 관계에 따른 목록화를 지원한다. 따라서 어휘 의미 구조와 관계에 대한 양적·질적 차원의 통합적 이해가 가능하다.

워드 임베딩은 단어의 의미가 주변 단어를 통해 결정된다는 분포 가설 (Distributional Hypothesis)(Sahlgren, 2008)을 전제로 특정 텍스트의 단어를 숫자로 표현하여 벡터화하는 자연어 처리 방법이다. 워드 임베딩은 통계적 방법으로 시작하여 최근에는 인공 신경망이나 딥러닝 방식을 활용한 학습 방식으로 발전해 왔다(임희석·고려대학교자연어처리연구실, 2020). 본 연구는 워드 임베딩의 학습 모델링 방법 중 신경망 구조를 활용하여 단어 간의 의미 유사도를 계산하는 Word2Vec(Mikolov, Chen, Corrado, & Dean, 2013)을 사용하였다.⁹⁾

워드 임베딩의 결과는 t-SNE를 활용하여 시각화하였다.¹⁰⁾ 고차원의 벡터를 2차원으로 축소하는 t-SNE(Maaten & Hinton, 2008)는 교과 어휘를 의미 유사도에 따라 2차원 공간에 배치하여 단어 간의 의미적 유사성을 유지하게 한다. 따라서 의미 거리가 가깝고 유사한 맥락에 분포하는 단어들은 의미적 군집을 형성하며, 이를 통해 텍스트 전체의 의미 구조를 보여준다. t-SNE 그래프는 Word2vec으로 모델링한 단어 중 빈도가 높은 단어와 단어 간 유사도가 높은 단어를 도출하여 구현하였다. 또 모든 학년군에 반복 출현하는 고빈도 공통 단어 50개를 그래프에 표시하여 학년에 따른 분포와 의미 군집의 변화를 확인하였다. 공통 단어의 목록은 다음과 같다.

-
- 9) Word2vec은 강형석·양장훈(2019)의 논의를 참고하여, Skip-gram 알고리즘을 적용하고, 벡터 크기 100, 윈도 크기 5, 최소 출현 빈도 3, 반복 학습 횟수 100으로 매개변수를 설정하였다.
- 10) t-SNE의 매개변수는 코퍼스 규모를 고려하여 n_components=2, perplexity=30, random_state=42로 설정하였다. 이는 Maaten & Hinton(2008: 2582)의 논의를 참고한 것이다.

지역, 사람, 사회, 문화, 문제, 생활, 모습, 경제, 다양, 국가, 도시, 활동, 환경, 고장, 해결, 이용, 세계, 변화, 인구, 일, 우리나라, 필요, 생산, 자료, 국제, 가족, 기후, 주민, 지도, 시장, 발생, 자원, 정보, 인간, 노력, 발달, 영향, 갈등, 조사, 인권, 가격, 역할, 개인, 자연, 과정, 권리, 국민, 상품, 정치, 사례

이들 단어는 사회과 교과서 전체 코퍼스에서 고빈도로 출현하면서, 동시에 다른 단어와의 관계에서 높은 벡터값을 갖는다.¹¹⁾ 사회과의 주제 영역의 핵심 개념을 표상하는 ‘문화’, ‘사회’, ‘지역’, ‘정치’, ‘경제’, ‘기후’ 등을 비롯하여 ‘과정’, ‘발생’, ‘사례’, ‘영향’과 같은 사고도구어, ‘모습’, ‘사람’, ‘일’, ‘가족’ 등의 기초 어휘를 포함한 추상성이 높은 단어였다.

네트워크 분석(network analysis)은 구성원 간 관계를 중심으로 패턴과 구조를 확인하여 시사점을 도출하는 사회과학적 분석 방법이다(Menczer, Fortunato, & Davis, 2016/2022). 언어 텍스트를 대상으로 한 네트워크 분석은 단어를 노드(node), 단어 간 연결을 링크(link)로 설정하여 의미망을 구성하고, 관계를 중심으로 패턴과 구조를 분석한다. 워드 임베딩이 단어 간 의미 거리를 통해 텍스트 전반의 의미 구조와 맥락을 추론한다면, 네트워크 분석은 단어의 관계에서 개별 단어의 영향력, 중요성, 의미화 양상을 미시적으로 확인하는 데 초점을 맞춘다.

본 연구는 학년군별 코퍼스 전체의 네트워크 분석을 선행하여¹²⁾ 사회과 교과서 어휘 네트워크의 전반적인 지표와 개별 단어의 중심성을 계산하는 한편, 그 하위 네트워크의 의미 관계의 양상을 살펴 네트워크(Ego Network)

11) 공통 단어 중에는 특정 학년에서 빈도가 낮지만, 벡터값은 높은 경우도 포함한다. 예를 들어 ‘국제’는 초등 3, 4학년군 코퍼스에서 18회 출현하는 저빈도 단어이나, ‘지도’, ‘이용’ 등의 단어와 0.7 이상의 벡터값을 갖고 있으며, 전체 학년군에서 총 1,551회 빈도로 출현 한다.

12) 학년군별 어휘 네트워크는 문장을 기본 단위로 한 동시 출현 기반 네트워크이며, 슬라이딩 윈도는 5로 설정하였다. 본 연구의 네트워크 분석은 의미 관계의 미시적 분석에 초점을 두었기 때문에 각 학년군별 전체 어휘 네트워크의 기본 지표는 생략하였다.

분석을 통해 확인하였다. 네트워크 기본 정보는 네트워크 규모를 확인할 수 있는 노드, 링크, 밀도, 평균 연결 수 등이 포함된다. 노드는 네트워크를 구성하는 단어 수이며, 링크는 노드 간 연결을 말한다. 밀도와 평균 연결 수는 네트워크 내의 활발하고 조밀한 연결을 설명한다.

중심성은 개별 노드가 전체 네트워크에서 수행하는 역할에 따른 중요도를 나타내는 지표이다. 본 연구에서는 연결중심성, 근접중심성, 매개중심성의 세 가지 지표를 살펴보았다. 연결중심성은 얼마나 많은 단어와 연결되는지, 근접중심성은 얼마나 짧은 경로로 연결되는지, 매개중심성은 해당 단어가 네트워크의 경로 중간에 얼마나 자주 등장하는지를 측정하여, 텍스트의 미 구성 과정에서 개발 단어의 역할을 확인하게 한다.

이고 네트워크(Ego Network) 분석은 ‘에고’인 핵심어와 그 이웃인 알터(alter)로 구성된 하위 네트워크를 통해 핵심어의 특성과 영향력을 파악하는 방법으로(이수상, 2012), 전체 네트워크 맥락에서 핵심어의 의미를 구체화하는 이웃 단어와 그 관계의 양상을 확인하는 것이다. 본 연구는 워드 임베딩 과정에서 검토한 공통 단어의 중 ‘국가’와 ‘시장’을 핵심어로 한 에고 네트워크를 생성하고 그래프로 시각화하였다. 또한, 이웃 노드를 ‘이전 학년중복 출현 단어’와 ‘해당 학년 최초 출현 단어’로 구분하여 학년군에 따른 핵심어의 의미화 양상 변화를 세부적으로 검토하였다.

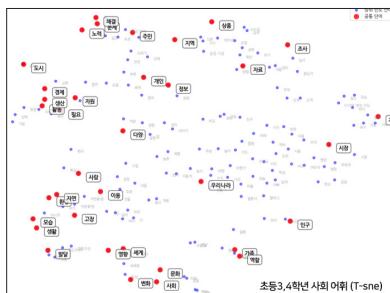
이상의 코퍼스의 정제 및 워드 임베딩, 네트워크 분석 과정은 Google Colab의 Python 3.10.12 버전에서 수행되었으며, gensim, networkx, scikit-learn, matplotlib 등의 라이브러리를 활용하였다. 또한 분석 대상 일반명사 중 실질적 의미가 약한 1음절 단어는 불용어로 처리하였다.¹³⁾

13) 사회과의 주요 교과 개념과 관련될 수 있는 ‘섬’, ‘강’, ‘산’, ‘길’, ‘쌀’, ‘차’, ‘봄’ 등은 불용어에 포함하지 않았다.

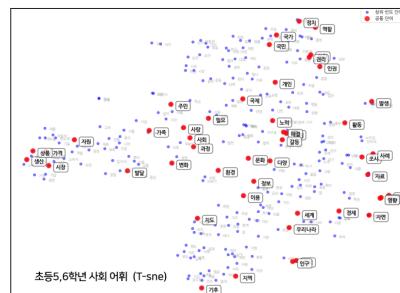
III. 연구 결과 및 분석

1. 워드 임베딩 결과 및 분석

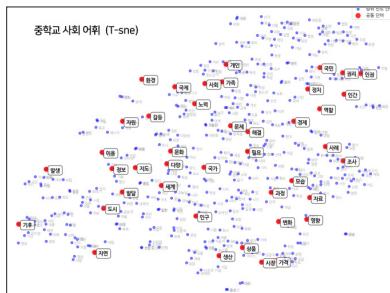
〈그림 1〉부터 〈그림 4〉는 학년군별 코퍼스의 워드 임베딩 결과를 t-SNE로 시각화한 그래프이다.¹⁴⁾



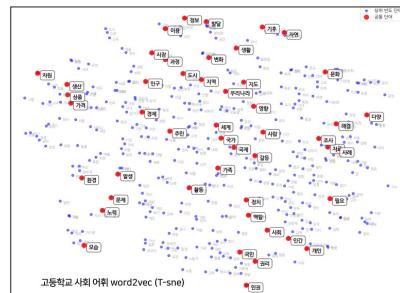
〈그림 1〉 초등3,4 t-SNE(169개)



〈그림 2〉 초등5,6 t-SNE(237개)



〈그림 3〉 중학교 t-SNE(423개)



〈그림 4〉 고등학교 t-SNE(374개)

14) 본 연구의 시각화 결과는 아래 링크에서 확대하여 확인할 수 있다. https://drive.google.com/drive/folders/17qM0LUUW44HvesIlAi3UUoqc64WGDPZg?usp=drive_link

위 그래프는 공통 단어의 맥락과 군집의 변화를 가시적으로 확인하기 위해 사회과 주요 주제와 관련이 낮은 단어를 제외하고, 학년군별로 어휘 유사도 0.7 이상, 상위 빈도 10% 해당하는 단어로만 간소화한 것이다.¹⁵⁾ 그림의 붉은 점은 앞서 제시한 공통 단어를 나타내며, 단어들은 의미의 상대적 거리에 따라 배치된다.

밀집된 단어 군집은 사회과의 다양한 주제를 반영한다. <그림 1>의 초등 3, 4학년 단어들은 넓게 분포하며 소수의 지역적인 군집을 형성한다. 군집 내 공통 단어를 볼 때, 이들은 경제, 지리, 사회 영역의 주제를 구성한다. 그러나 이들 영역과 의미적으로 밀접한 ‘시장’, ‘인구’, ‘지도’ 등이 해당 영역과 거리를 두고 있어, 군집 내 단어들은 각 주제 영역의 개별 맥락에서 사용되었을 것으로 보인다.

초등 5, 6학년 이후 단어들은 <그림 2>에서 보듯, 밀접도가 높아지고, 군집 규모가 확대되며 그 거리도 좁아진다. ‘상품’, ‘생산’, ‘자원’ 등의 경제 영역의 군집에 ‘시장’, ‘가격’이 추가되었으며, ‘가족’, ‘사람’, ‘사회’ 등이 포함된 일반 사회 영역 군집과의 거리는 이전 학년보다 가까워졌다. 또한 우측 상단에는 정치 영역의 군집이 새롭게 형성되었다. 다만 명확히 분리된 군집은 아니며, 주변부에서 일부 단어를 매개로 다른 영역과의 연계가 이루어지는 것을 볼 수 있다.

초등학교와 중·고등학교 어휘의 의미 분포와 군집의 차이는 확연하다. <그림 3>, <그림 4>의 중·고등학교 단어들은 간격이 좁고 비교적 군집 하며, 독립된 군집을 구별하기도 어렵다. 이는 사회과 어휘의 의미 군집이 개별적 의미에서 통합적으로 변화하고 있음을 보여주며 학년이 올라갈수

15) 워드 임베딩의 모델링에 포함된 단어는 초등 3, 4학년 1,652개, 초등 5, 6학년 2,373개, 중학교 4,117개, 고등학교 3,714개이다. 각 코퍼스의 규모가 다르므로 동일한 분량의 단어를 표시하지 않고, 일정 비율 내의 단어가 그래프에 반영되게 조정한 것이다. 학년군별 그림은 차원이 축소된 형태지만, 어휘의 양적 규모를 반영하여 학년에 따른 단어 증가 추이도 짐작할 수 있다.

록 단어 의미도 개별 맥락이 아닌 사회과 전체의 확장된 맥락에서 결정됨을 의미한다.

지엽적으로 분포하던 공통 단어가 학년이 올라가며 군집 전체에 두루 분포하는 양상도 이러한 해석을 뒷받침한다. 특히 사회과 주제를 나타내는 ‘경제’, ‘문화’, ‘국가’, ‘사회’, ‘정치’ 등의 단어는 시기별로 분포 변화가 활발하다. 초등 시기에 이들 단어는 주로 주변부의 개별 군집에 분포하지만 중·고등학교에서는 대부분 중심에 가깝게 분포하며 다양한 영역의 단어들과 의미 거리를 좁힌다.¹⁶⁾ 반면 ‘가격’, ‘생산’, ‘시장’, ‘인권’, ‘권리’, ‘기후’ 등은 <그림 1>~<그림 4>에서 학년이 올라가도 여전히 주변부에 머문다. ‘사회’, ‘문화’, ‘국가’나 ‘시장’, ‘인권’, ‘기후’는 모두 사회과 교과 지식 개념을 표상하는 어휘이지만, 초등 고학년 이후 사회과의 의미 구조가 점차 연계되고 통합되면서 이들 어휘의 의미 기능은 차별화되는 것으로 보인다. ‘사회’, ‘국가’가 여러 영역 간의 통합과 연계를 담당하며 사회과 전체 의미 구성에 기여하는 반면, ‘인권’이나 ‘시장’ 등은 개별 영역의 전문성과 특수성을 드러내어 교과 지식의 심화를 담당한다.

이상의 워드 임베딩의 결과는 학년에 따라 변화하는 사회과 어휘의 의미 구조의 양상을 보여준다. 사회과 교과를 구성하는 어휘는 초등 고학년 이후 주제 영역이 구체화되면서도 연계와 통합도 이루어진다. 그 의미가 결정되는 맥락도 다양해지며, 교과 어휘의 의미 기능의 차이도 분명해진다. 통합과 연계를 담당하며 교과 전체의 의미적 구조에 기여하는 어휘와 주제 영역의 특수성을 드러내는 어휘 등, 교과 어휘의 의미 기능의 차이가 분명해지는 변화의 과정도 확인할 수 있다.

16) t-SNE의 중심부에 가까운 단어가 텍스트 전체의 주요한 핵심 단어이거나 의미상 중요한 단어를 나타내는 것은 아니다. 그러나 t-SNE가 단어 간 의미 거리를 반영한다는 점을 고려하면 중심부의 단어는 다양한 의미 맥락의 단어와 연결된 가능성이 높으며, 주변부의 단어는 상대적으로 의미 유사성이 높은 단어와의 연결이 제한적일 수 있다.

2. 네트워크 결과 및 분석

이 절에서는 워드 임베딩의 결과를 바탕으로 학년에 따라 의미 위상의 차이가 두드러진 ‘국가’와 ‘시장’을 핵심어로 하는 에고 네트워크 분석을 수행하고, 그 의미화 양상을 주변 단어와의 관계를 통해 확인하였다. ‘국가’는 학년이 올라갈수록 의미 맥락이 확장되었으나, ‘시장’은 학년이 올라가도 여전히 주변부의 제한된 의미 맥락에 머무는 단어였다.

표준국어대사전에서 ‘국가’는 “일정한 영토와 거기에 사는 사람들로 구성되고, 주권(主權)에 의한 하나의 통치 조직을 가지고 있는 사회 집단. 국민·영토·주권의 삼요소를 필요로 한다”로 정의되는 단의어이며, ‘시장(市場)’은 “여러 가지 상품을 사고파는 일정한 장소”, “상품으로서의 재화와 서비스의 거래가 이루어지는 추상적인 영역”이라는 두 가지 뜻이 있는 다의어이다. ‘국가’와 ‘시장’을 핵심어로 하는 에고 네트워크 분석에서는 단의어이지만 의미 맥락이 확장적인 ‘국가’와, 다의어임에도 제한적인 맥락에서 의미화되는 ‘시장’의 의미적 특성에 주목하여 각 의미 관계의 변화를 검토하였다.

‘국가’와 ‘시장’의 학년별 에고 네트워크의 기본 정보와 전체 네트워크 내에서 두 단어의 중심성 지수는 다음 〈표 2〉와 같다.

〈표 2〉 ‘국가’와 ‘시장’의 에고 네트워크 규모와 중심성 지수

구분	국가				시장			
	E34	E56	M	H	E34	E45	M	H
노드	27	170	641	474	72	74	265	153
링크	145	2536	17154	10741	623	674	5617	2502
밀도	0.041	0.177	0.084	0.096	0.244	0.250	0.161	0.215
평균 연결 수	10.74	29.83	53.52	45.32	17.31	18.22	42.39	32.71
연결 중심성	0.010	0.057	0.141	0.102	0.033	0.019	0.051	0.033
근접 중심성	0.242	0.319	0.365	0.312	0.295	0.280	0.312	0.261
매개 중심성	0.0002	0.013	0.026	0.020	0.005	0.002	0.008	0.004

두 단어의 에고 네트워크 규모와 중심성은 학년이 올라가면서 중학교까지 대체로 증가한다.¹⁷⁾ 노드와 링크 수, 평균 연결 수, 중심성 지수의 증가는 네트워크 전체 규모의 확장을 의미한다. 반면, 연결의 조밀도를 의미하는 밀도는 학년이 올라갈수록 낮아지는데,¹⁸⁾ 각 네트워크에서 노드와 링크가 증가하면서 특정 노드에 연결이 집중되지 않고 고르게 확산된 결과로 보인다. 즉, 밀도는 낮아졌지만 연결 구조는 더 다양해지고 복잡해진 것이다.

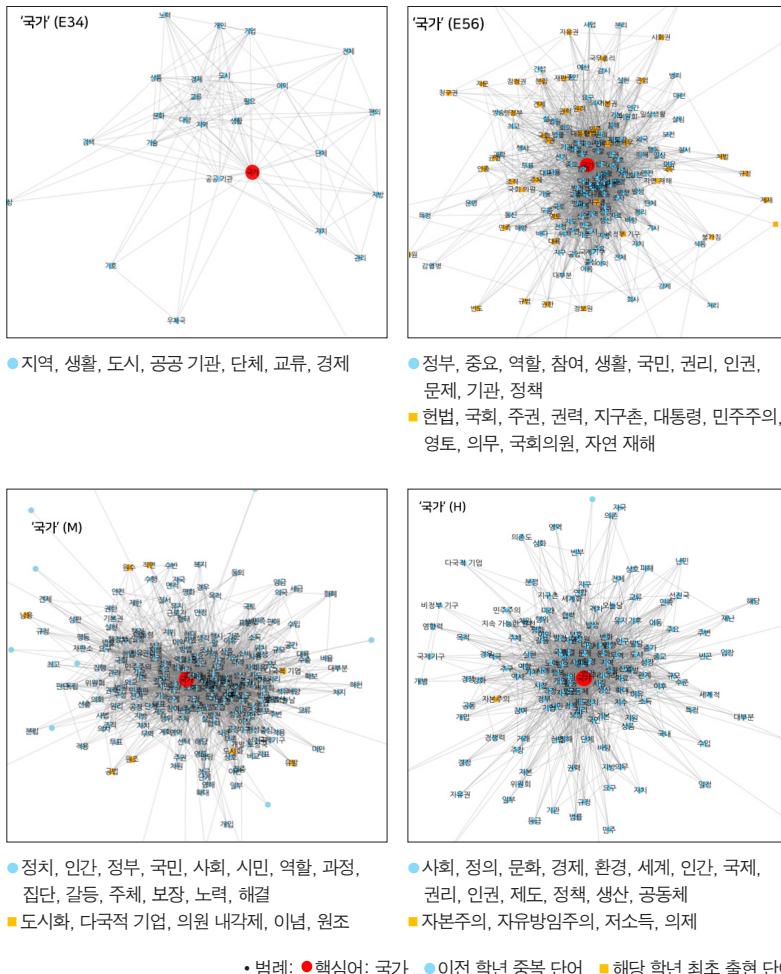
중심성 값은 두 단어의 역할과 의미 기능을 간접적으로 보여준다. 초등 학교 5, 6학년 이후부터 ‘국가’의 중심성 값은 ‘시장’보다 모든 항목에서 높았다. 중심성이 높은 단어가 텍스트의 주제나 핵심 개념을 나타내는 경우가 많다는 점(Menczer et al., 2016/2022: 104)을 고려하면 ‘국가’가 ‘시장’보다 사회과 전체의 주제 의미를 구성하는 주요 키워드가 될 수 있음을 보여준다.

두 단어의 네트워크나 중심성 값의 변화에도 차이가 나타났다. ‘국가’는 초등 3, 4학년에서 네트워크의 규모와 중심성 값이 ‘시장’보다 낮았으나, 5, 6학년부터는 더 높은 값을 보였다. ‘시장’은 초등 3, 4학년과 5, 6학년의 차이가 크지 않고, 오히려 5, 6학년에서 값이 낮아지기도 했다. ‘시장’의 중심성은 중학교에서 이전 학년에 비해 증가하였으나, 그 값은 ‘국가’보다 작았다.

이러한 양상은 두 단어의 에고 네트워크에서도 확인된다. 〈그림 5〉, 〈그림 6〉은 ‘국가’와 ‘시장’의 학년별 에고 네트워크의 결과이다. 그래프의 링크 길이는 에고와 이웃 노드의 연결 수에 반비례하도록 조절하였다.¹⁹⁾ 즉, 붉은 원으로 표시된 핵심어와 가까운 이웃 노드는 두 단어와 연결 빈도가 높은 단

-
- 17) 두 단어 모두 고등학교에서 에고 네트워크 규모와 중심성이 감소하는데, 이는 고등학교 코퍼스의 규모가 중학교보다 작기 때문이다.
 - 18) 네트워크 밀도의 계산은 ‘밀도(D) = 실제 링크 수 / 최대 가능 링크 수’에 따라 계산된다 (Menczer et al., 2016/2022.). ‘최대 가능 링크’는 연결 가능한 모든 경우의 수이므로 노드 수에 비례한다.
 - 19) 그래프는 Spring Layout으로 설정하고, 식별을 위해 연결 수가 적은 단어를 필터링하였다. E34의 네트워크에서는 링크 수 3 이하, 그 외의 네트워크는 5 이하 단어를 제외하도록 매개변수를 조정하였다.

어들이다. 이웃 노드는 이전 학년군에서 중복 출현한 단어와 해당 학년에 최초 출현한 단어로 구분하였으며(그림 하단의 범례 참조), 에고와 가까운 일부 단어를 그래프 하단에 목록으로 제시하였다.



〈그림 5〉 ‘국가’의 학년군별 에고 네트워크

〈그림 5〉에 따르면, ‘국가’의 개념이 구체화되는 시기는 5, 6학년이다. 단순한 의미 맥락을 구성하던 3, 4학년의 네트워크와 달리 5, 6학년에는 새로운 단어가 다수 도입되며 의미망이 확장된다. 중·고등학교에는 새로운 단어의 도입보다는 이전 단어와의 활발한 연결을 통해 복잡하고 확대된 의미망을 형성하고 있었다.

‘국가’의 이웃 단어의 양상은 다면어로서²⁰⁾ 의미적 특징을 반영한다. 초등 3, 4학년에서 ‘국가’는 ‘공공 기관’과 가장 가깝게 연결된다. 초등학교 4학년 1학기에 ‘공공 기관’을 정의하며 출현하는 ‘국가’는²¹⁾ 일상과 밀접한 공간을 관리하는 〈집단〉으로 제시되며, ‘지역’, ‘도시’와 연결을 통해 〈공간〉의 국면을, ‘경제’, ‘교류’를 통해 〈역할과 기능〉의 국면도 일부 드러내고 있었다.

초등학교 5, 6학년의 ‘국가’는 정치 영역의 주제를 본격적으로 다루며 그 의미 국면이 확장된다. 이전 학년에서 출현한 ‘정부’, ‘국민’과 함께, 새로 도입된 ‘헌법’, ‘국회’, ‘대통령’, ‘영토’, ‘참여’, ‘권리’, ‘민주주의’ 등은 ‘국가’의 〈기관〉, 〈원리〉, 〈주체〉, 〈집단〉, 〈역할과 기능〉의 의미 국면을 부각한다. 이러한 의미 국면은 정치 영역의 지식 체계를 반영한 것이며, 앞서 확인한 ‘국가’의 사전적 의미를 대부분 포괄한다. 따라서 ‘국가’의 의미는 이러한 다양한 국면의 통합을 통해 구성된다. 사회과 교과서에는 ‘국가’의 개념이 명

20) 임지룡(2018: 141)에 따르면 다면어(多面語)는 그 의미가 둘에서 네 개 정도의 ‘국면’으로 구성되며, 각 국면은 한 단어의 의미를 상호 보완적으로 연결하여 ‘통일체(gestalt)’를 형성하는 단위이다. 그는 ‘국가’를 ‘영토’, ‘국민’, ‘주권’의 세 가지 국면을 가진 다면어로 소개한다. 임지룡(2018)은 ‘다면어’가 ‘다의어’나 ‘단의어’와는 다른 의미 유형이라 설명 하지만, 이를 단의어나, 다의어의 일종으로 보는 연구도 있다(안해지, 2018). 본 연구는 단어의 범주에 구애되지 않고 의미의 다양성을 설명하는 관점으로 ‘다면어’의 개념을 전제하여 ‘국가’를 분석하였다.

21) ‘공공 기관’은 다음과 같이 정의된다 “지역에 있는 다양한 장소 중에서 개인의 이익이 아닌 주민 전체의 이익을 위해 국가나 지방 자치 단체가 세워 관리하는 곳을 공공 기관이라고 합니다”(김정인·박승규·박윤경·정요근·이종원·김자영 외, 2022). 99, 밑줄은 연구자)

시적으로 정의되지는 않지만, ‘국가’의 다양한 의미 국면은 해당 주제 단원의 학습 요소이며, 이들의 학습은 ‘국가’의 의미를 통합적으로 이해하는 출발이 된다.

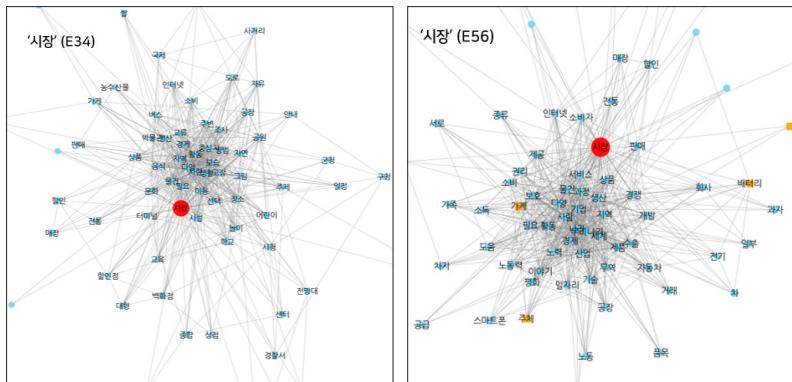
중학교 이후 ‘국가’는 기존 단어와의 연결을 통해 그 의미 맥락이 확장된다. 특히 ‘사회’, ‘문화’, 경제 등 사회과의 여러 주제 영역을 표상하는 단어와의 연결은 ‘국가’의 의미 맥락을 다른 주제 영역으로 확장하는 양상을 보여준다. 이는 앞서 확인한 워드 임베딩 결과와도 일치된 것이다. 이때 ‘국가’는 의미 국면의 일부를 활성화하며 서로 다른 영역을 연결하고, 새로운 개념을 구체화하는 역할을 한다. 이는 중학교 이후 네트워크에 새로 추가된 단어인 ‘도시화’, ‘자본주의’ 등을 통해 확인할 수 있다. 다음은 중학교 사회 교과서에서 ‘도시화’와 ‘국가’를 포함한 문장이다.

도시화 단계는 경제 성장과 밀접한 관련이 있기 때문에 이를 통해 한 국가의 경제 수준을 파악할 수 있다.

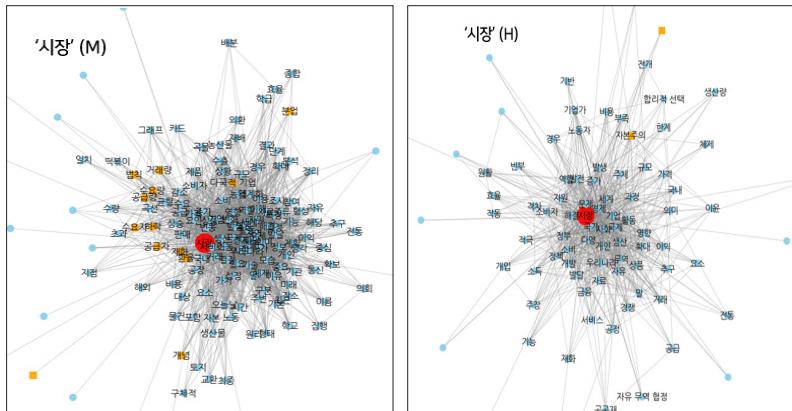
(최성길·최원희·강창숙·박상준·최병천·조일현 외, 2018: 151, 밑줄은 연구자)

‘도시화’는 지리 영역의 개념이지만²²⁾ 예문은 ‘도시화’를 경제의 문제와 연결한다. 이러한 연결을 위해서는 ‘국가’의 다양한 의미 국면을 중 경제 주체로서 의미가 부각되어야 한다. 이 과정에서 다른 국면이 배제되는 것은 아니기 때문에, ‘국가’의 또 다른 국면은 ‘도시화’를 사회과의 다른 영역과 연결하기도 한다. 따라서 ‘국가’의 의미는 맥락에 따라 달라지는 것이 아니라, 부각되는 의미의 국면이 달라지는 것이며, 학습은 이러한 다양한 의미의 국면을 발견하고 체계화하는 과정이 된다.

22) 위 문장은 도시화의 부가적인 정보를 기술하며, 교과서에서는 이를 전제로 ‘선진국’, ‘개발도상국’의 도시화를 비교하는 활동이 이어진다.



- 시설, 선택, 이용, 장소, 물건, 문화, 고장, 활동, 중심지, 지역, 음식, 상품
- 소비자, 서비스, 상품, 판매, 전통, 생산, 과정, 제공, 매장, 기업
- 가계, 주체, 배터리, 주식



- 서비스, 생산, 판매, 거래, 국내, 변동, 상승, 공급, 물가, 가격, 요인, 결정, 필요, 소비, 농업, 영향, 감소
- 환율, 재화, 공급자, 수요자, 하락, 공급량, 법칙, 거래량, 다국적 기업, 법칙, 분업
- 경제, 국가, 문제, 해결, 지원, 기업, 활동, 생산, 개인, 정부, 소비, 자원, 증가, 발생
- 자본주의, 외부 효과

• 범례: ● 핵심어: 시장, ● 이전 학년 중복 단어, ■ 해당 학년 최초 출현 단어

〈그림 6〉 '시장'의 학년군별 애고 네트워크

한편, 다의어인 '시장'의 의미 관계의 변화는 '국가'와는 다른 양상을 보인다. 〈그림 6〉에서 보듯 초등 두 개 학년군의 '시장'의 네트워크 규모는 큰

차이가 나타나지 않았으며, 중학교에서 가장 복잡한 연결 양상을 보였다. 새로운 단어의 도입도 중학교 시기에서 가장 활발했다.

초등학교 6학년은 경제 영역이 초등 과정에서 처음 다루어지며 ‘시장’의 개념을 본격적으로 배우지만, 5, 6학년의 네트워크 규모가 다른 학년군에 비해 작았다. 새로운 단어의 도입도 4개에 불과한 것은 예상과는 다른 결과였다. <그림 6>에서 ‘시장’과 연결된 이웃 단어는 이러한 결과가 ‘시장’의 학년군별로 다른 의미 기능의 차이, 다의어로서의 특징, ‘시장’의 의미화 과정을 반영한 것임을 보여준다.

초등 3, 4학년에서 ‘시장’은 앞서 확인한 사전의 첫 번째 의미인 <공간>이다. ‘장소’, ‘중심지’ 등으로 개념화되는 ‘시장’은 일상의 경험을 통해 습득될 수 있는 단어이다. 아래는 ‘시장’이 처음 출현하는 3학년 사회 교과서의 문장이다.

우리 고장에는 집, 학교, 놀이터, 시장, 산, 공원 등 다양한 장소가 있습니다.

(김정인 · 박승규 · 박윤경 · 정요근 · 이종원 · 김자영 외, 2022 그: 12, 밑줄은 연구자)

여기서 ‘시장’은 경제 영역이 아닌 지리 영역에서 ‘장소’, ‘공간’, ‘지역’의 의미를 구체화하기 위한 역할을 담당한다. <표 2>에서 3, 4학년의 ‘시장’의 매개 중심성이 상대적으로 높게 나타난 것은 이러한 ‘시장’의 의미 기능이 반영된 것이라 할 수 있다.

6학년 1학기에서 ‘시장’은 경제 단원의 지식 개념으로, 사전의 두 번째 의미인 <추상적인 영역>으로 나타난다. 3, 4학년의 ‘시장’이 중심의미라면, 이 시기의 의미는 주변의미에 해당한다. 다의어는 구체적이고, 지각 가능한 중심의미에서 추상적이고 심리적인 의미로 확장되며 생성되는데, ‘시장’처럼 공간이 추상적인 개념으로 확장되는 것은 가장 빈번하고 일반적인 의미 확장의 경로라 설명된다(임지룡, 2018: 136). 특히 ‘시장’의 두 의미는 ‘사고 팔다’라는 유사성을 중심으로 의미가 확장된 것이며, 이러한 은유적 기제에

의한 의미화는 사회나 과학 등의 전문 지식 개념을 표상하는 방식으로도 알려져 있다. 사회과 교과서는 이러한 의미화 방식을 ‘시장’을 설명하는 방식에도 활용한다.

시장은 물건이나 서비스를 사려는 사람과 팔려는 사람이 모여 거래하는 곳이다.
할인 매장이나 인터넷 쇼핑몰처럼 물건을 사고파는 시장도 있지만, 여러 나라의 돈이나 회사의 일부를 소유할 수 있는 권리, 집이나 땅을 사고파는 시장도 있다.

(김정인·박승규·박윤경·정요근·이종원·김자영 외, 2023: 81, 밑줄은 연구자)

주변의미로서 ‘시장’은 신체적으로는 지각할 수 없는 추상적 의미이지만 ‘거래’, ‘사고팔다’라는 공통의 개념을 활용하여 의미를 구체화하며, 공간으로서 ‘시장’보다 경제 활동에 관련된 더 많은 현상을 설명할 수 있게 되는 것이다.

5, 6학년 네트워크에서 보이는 ‘시장’의 이웃 단어는 대부분 이전 학년 군에서 <공간>으로서 ‘시장’을 설명한 단어이며, 새로 도입된 단어는 ‘가계, 주체, 배터리, 주식’에 불과하다. 이는 유사성에 근거한 비유적 설명의 방식을 활용하여 주변의미인 ‘시장’을 설명하기 때문으로 보인다.

중학교 이후, 공간으로 ‘시장’을 의미화하는 연결은 거의 보이지 않는다. ‘시장’과 연결이 활발한 이웃 단어들은 주로 추상적인 개념으로, ‘환율’, ‘재화’, ‘공급자’, ‘수요’ 등의 경제 영역의 전문 개념을 표상하는 다수의 단어가 도입되는 것을 볼 수 있다. 중학교 이후 확장된 의미로서 ‘시장’은 더 이상 직접적 교육 내용 요소가 아니며, 보다 전문적인 개념과의 연결을 통해 경제 분야의 지식을 심화하는 역할로 전환되었음을 보여주는 것이다.

‘시장’이 타 영역으로의 확장되는 것은 확인하기 어려웠다. 고등학교 이후 ‘시장’은 ‘경제’, ‘국가’ 등 개념의 범위가 넓은 기준 단어와 활발히 연결되지만, 새로 도입된 단어는 ‘자본주의’와 ‘외부 효과’가 전부였다. <추상적 영역>으로서 ‘시장’의 의미는 ‘시장’이 갖는 <공간>, <기능>, <구성원> 등의 의미

국면 중 <기능>인 ‘사고팔다’만 부각되고 나머지는 배제된다. 고등학교의 사회과 전체 의미 구조가 영역 간 연계와 통합이 이루어지는 방향으로 변화하더라도, ‘시장’의 의미는 한정적이기 때문에 ‘국가’만큼 타 영역과의 연계가 활발하게 이루어지지 않았던 것이다.

요컨대, ‘국가’와 ‘시장’의 학년별 애고 네트워크는 교과 지식 체계에서 교과 어휘의 의미 구성 방식과 의미 기능의 세부적 변화 및 상호 관계를 보여준다. ‘국가’는 다면어로서 다양한 의미 국면의 통합을 통해 그 의미가 구성되었으며, ‘시장’은 다의어로서 구체적 공간을 추상적 영역으로 확대하며 의미가 구성되었다. 이러한 차이는 맥락에 따라 의미가 구체화되는 방식뿐만 아니라, 교과 지식 개념이 제시되는 방식과 의미 기능이 변화되는 양상과도 관련됨을 확인할 수 있었다. 이는 개별 단어의 의미가 교과 전체의 의미 구성에 기여할 수 있음을 보여준다.

IV. 논의와 시사점

사회과 교과서 어휘의 워드 임베딩과 네트워크 분석 결과는 새로운 지식을 구성하고 확장하는 매개로서의 교과 어휘의 역할을 보여준다. 동일한 단어라도 학습의 과정과 시기에 따라 의미 기능이 달라지며, 이는 의미가 언어 자체에 있는 것이 아니라 맥락과 관계 속에서 구성되는 것임을 확인해 준다. 의미가 맥락에 의해 구성되고 결정되는 것이라면, 교과 어휘, 특히 교과 지식 개념을 표상하는 어휘의 맥락은 교과가 터한 학문 및 사회 공동체의 지식 소통 방식과 관습이다. 지식은 언제나 상대적이며, 사회적으로 공증된 결과물로서 그것을 구성한 구성원과 상호작용 속에서 가치를 부여받기 때문이다(장상호, 1997: 50).

비록 사회과라는 한정된 교과의 사례이지만, ‘국가’와 ‘시장’은 이러한

의미와 맥락의 상호작용을 보여준다. 이를 단어의 의미화 방식은 교과서 텍스트 안에서 구성하고 있는 의미 구조와 의미 기능에 반영되며, 그 의미 네트워크는 교과 전체 텍스트에서 두 단어의 의미 기능의 차이를 보여주었다. 따라서 교과 교육은 학문적 지식 개념이 소통되는 다양한 맥락을 제공하는 장이며, 교과의 지식을 이해하는 학습은 개념을 표상한 언어의 축자적 의미를 배우는 과정이 아니라 개념을 둘러싼 맥락을 이해하고 체계화하여 의미를 찾고 발견하는 과정이라 할 수 있다.

이러한 관점에서 보면, 학습을 통해 어휘가 발달하는 것은 의미에 대한 이해가 확장되고 심화되는 것이다. 교과가 제공하는 다양한 맥락이 추상적이고 전문적인 맥락으로 확장되며, 학습자는 이를 매개하는 언어의 의미를 발견하기 위해 단어의 의미 기능을 인식하고 다양한 방법으로 의미를 구성해야 한다. 따라서 추상적인 언어, 은유적인 언어 등으로 설명되는 교과 어휘는 그 의미 구성의 결과물이다. 언어 발달 연구에서 설명하는 학령기 아동 및 청소년의 추상적 언어, 전문적 언어의 확장 역시 특정 단어군에 대한 소유나 양적 확장으로 볼 것이 아니라, 해당 단어가 표상하는 의미의 복잡성, 관계성을 이해하고 다루는 능력의 확장으로 보아야 하는 것이다.

앞서 살펴본 ‘국가’나 ‘시장’은 이러한 의미 발달의 과정을 통해 교과 지식이 구성되는 사례를 보여준다. 초등 3, 4학년 교과서에 처음 등장하는 ‘국가’와 ‘시장’은 학교 이전에도 접할 수 있지만, 추상적이거나 전문적인 개념으로 이해되기는 어렵다. 이를 단어는 초등 고학년이 되어서야 사전적 의미에 가까운 내용이 학습 내용으로 제시되며, 이때 ‘국가’는 다면적 의미의 국면을 통합적으로 이해해야 하며, ‘시장’은 중심의미와 확장된 의미를 연결하여 초점화하는 능력이 요구된다. 중학교 이후 이러한 추상적 개념은 다른 단어와의 연결을 통해 새로운 개념 형성의 기반이 되므로, 이 시기 ‘국가’와 ‘시장’에 대한 얇은 단순한 개념 습득을 넘어, 그 기능에 대한 인식도 포함한다고 할 수 있다.

결국 추상적이고 전문적인 의미로서 ‘국가’와 ‘시장’이 수용되기 위해

서는 일상의 언어와는 다른 교과 어휘의 의미 특성과 기능, 언어의 다의성과 다면성에 대한 인식과 이해가 전제되어야 한다. 또한 교과 지식 맥락에 근거 하여 의미를 발견하고 확장함으로써 새로운 의미를 구성할 수 있어야 한다.

다른 교과 역시 세부적 양상은 다르겠지만, 어휘로 인한 문식성의 문제는 결국 의미 구성의 문제이다. 새로운 개념을 표상하는 어휘가 추가되는 것과 더불어 친숙했던 단어라도 학년이 올라갈수록 그 개념적 구조와 관계가 확대되고 복잡해지면 그것이 사용되는 순간에 요구되는 의미를 포착하는 일은 쉽지 않다. 따라서 학습을 통한 교과 어휘의 발달은 의미를 둘러싼 복잡한 맥락에서 의미를 결정하는 능력을 기르는 과정이며, 이는 의미 기능의 인식과 의미 구성 능력의 점진적 변화라는 두 가지 양상으로 설명될 수 있을 것이다.

결과로서 어휘가 아닌 의미의 발달과 성장으로 교과 어휘를 바라보면, 기준과는 다른 교과 어휘의 교육 내용과 방법이 요구된다. 더 구체적이고 본격적인 논의가 필요하겠지만, 본 연구를 통해 제안할 수 있는 교과 어휘 교육의 방향은 다음과 같다.

첫째, 어휘 교육의 대상으로서 교과 어휘는 제재로서 제시되어야 한다. 국어과 어휘 교육의 대상으로 사고도구이나 교과 개념어를 제시한 신명선(2004)과 전영주(2019)는 이를 어휘가 개방 집합임을 강조한다. 이 연구들은 학술 텍스트나 교과서 텍스트에 포함된 어휘를 빈도, 분포, 의미 기능 등을 통해 분류하고 목록화한 결과를 제시했지만, 이러한 목록은 최소한의 결과일 뿐이다. 실제 언어 사용의 맥락에서는 더 많은 단어가 목록에 추가될 수도 있으며, 개별 단어의 의미 기능도 달라질 수 있다.

따라서 어휘 교육의 제재로 활용할 어휘를 학습자에게 제공할 때는 빈도나 분포보다는 특징과 의미 기능, 즉 텍스트의 의미 구성에 기여하는 역할을 우선 고려해야 한다. 특히 교과 어휘는 교과 지식 구성 과정에서 요구되는 의미의 다양성, 복잡성을 경험할 수 있는 단어가 사례로서 제시될 수 있다.

이러한 관점에서, 일상에서 만나기 어려운 낯선 어휘뿐만 아니라 일상

의 단어지만 의미가 확장되어 전문적 개념을 갖게 된 어휘, 학문적 체계화를 통해 개념화된 어휘(예: 기하급수, 스펙트럼) 등도 교과 어휘 교육의 제재가 될 수 있다. 또한, 이미 익숙한 일상의 단어 역시 타 교과 및 다양한 주제 영역의 맥락에서 새롭게 의미화되는 경험을 제공할 수 있다면 어휘 교육의 제재로서 충분히 활용될 수 있을 것이다.

둘째, 교과 어휘 교육을 위한 내용은 개별 단어의 의미 자체가 아닌 의미 기능과 의미 구성을 위한 방법이 되어야 한다. 사전이 제공하는 의미는 다의어라고 할지라도 이미 결정된, 공인된 결과로서의 의미가 등재된 것이다(정한데로, 2023: 221 - 223). 사전은 객관적이고 잠정적인 의미를 파악할 수 있다는 점에서 참고될 수 있으나 그 자체가 어휘 교육의 종착점이 될 수는 없다. 특히 교과 어휘는 교과의 구체적인 맥락에서 그 의미가 결정되며, 여러 영역의 주제를 넘나들며 연계와 통합, 심화가 이루어진다. 따라서 사전적 의미를 출발점으로 삼되, 의미 기능과 의미화 방식을 적용하는 과정을 중심으로 교육 내용이 구성되어야 한다.

이때 교과 어휘의 의미 기능은 사고도구어나 교과 개념어를 구분하는 기준처럼 주제의 일반성과 특수성에 따라 구분되거나, 구성하는 지식의 역할에 따라 구분될 수도 있다.²³⁾ 또한 의미화 방식은 학습자의 인지 발달과 문식성 발달을 고려해 단계적으로 구체화해야 하며, 이 과정에서 인지언어학의 연구 성과와 타 교과의 지식 구성 관습의 특수성도 반영할 수 있을 것이다. 사회나 과학, 혹은 각각의 하위 주제 영역은 분야마다 명명화 방식이나 의미화 방식에서 차이를 보일 수 있기 때문이다. 결국 이러한 구분은 어휘의 의미를 결정하고 구성하기 위해 인식해야 할 맥락들로, 교과 어휘 교육에서는 일상의 언어와는 다른 교과 어휘의 특수성에 대한 인식과 이해, 그리고

23) 예컨대, 사고도구어의 의미 기능적 역할에서 설명하는 '수행어', '대상어'(신명선, 2004: 152) '조직어'는 각각 '절차적 지식', '개념적 지식', '조건적 지식'을 표상하는 어휘로서 구분될 수 있을 것이다.

의미의 역동성과 다양성에 대한 인식과 이해도 포함되어야 할 것이다.

셋째, 교과 어휘 교육은 국어과 문식성 교육과의 연계를 고려하여 직접적인 교육 방식으로 구체화되어야 한다. 개별 단어의 의미를 문맥에서 찾고, 그 사전적 의미를 확인하는 방식은 가장 일반적인 어휘 교육의 방법이었다. 그러나 학년이 올라갈수록 교과의 의미 구성 과정이 복잡한 맥락과 확장적인 관계를 통해 이루어지는 것을 고려하면 짐작과 사전을 통해 이루어지는 어휘 학습이 교과 학습의 의미 구성 과정에 기여하는 일은 한계가 있을 수밖에 없다. 따라서 문맥에서 의미를 짐작하는 활동은 적확한 의미를 찾는 것에서부터 교과의 특수한 상황과 텍스트 이면의 저자 및 학문 공동체의 의도까지 탐구하는 활동으로 확장되어야 한다. 단어의 의미 관계를 파악하는 활동도 개별 단어 수준이 아닌 텍스트 차원의 유사한 의미, 확장적 의미, 상반된 의미를 확인하며 의미를 정교화하고 새롭게 구성할 수 있는 방향으로 발전되어야 한다.

그간 국어과 기능 영역에서는 의미 구성을 위한 학습자의 인지적 기능과 전략을 강조하며, 어휘를 문식성 교육의 부차적인 요소로 다루어 왔다(전영주, 2020). 이는 사고 활동으로서 언어 수행을 강조한 결론이기도 하지만, 언어 수행 과정에서의 어휘 역할에 대한 연구나 관심이 부족했기 때문이기도 하다. 도구 교과로서 국어과 교육이 타 교과 학습에 기반이 되는 어휘 교육을 제공하려면, 국어과 기능 영역에서 어휘의 역할을 구체화하여 어휘를 통해 의미를 구성하는 직접적 어휘 교육의 내용과 방법이 모색되어야 할 것이다.

V. 결론

단어의 의미는 본질적으로 그 경계가 불명확하다(임지룡, 2018: 126). 따라서 교과의 학습은 단어 자체의 의미를 획득하는 과정이 아니라 불명확한

경계를 명확하게 하는 과정이라 할 수 있다. 교과 학습이 불명확한 경계를 명확히 하는 교과의 특수한 맥락과 체계를 제공한다면, 국어과 교육은 이를 매개하는 언어의 역할과 의미를 발견하는 방법을 안내하는 역할을 한다.

본 논의는 이러한 국어과의 역할을 고민하며 사회과 교과 어휘의 의미 구조와 관계 변화를 탐색하였다. 워드 임베딩과 네트워크 분석을 통해 이를 확인한 것은 텍스트 전체 구조와 개별 단어의 의미가 긴밀하게 연결되며 의미와 맥락이 상호작용 한다는 것을 전제한 것이다. 또 이를 학년군별로 비교하여 그 변화 과정을 살펴본 것은 어휘 발달의 양상을 통해 어휘 교육의 방향과 내용을 고민하기 위한 시도였다.

사회과라는 한정된 교과를 대상으로 한 연구였지만, 학습을 통해 어휘가 발달한다는 것은 결국 의미가 발달하는 것이며, 어휘의 습득은 그 결과임을 확인할 수 있었다. 이러한 논의는 문식성의 기반이 되는 교과 어휘를 결과로서의 어휘가 아닌 의미 발달의 과정으로 인식할 필요가 있음을 보여준다. 또한 국어과의 어휘 교육 역시 단어의 의미 학습을 넘어, 의미를 발견하고 새롭게 구성하는 기능과 방법의 차원에서 재고되어야 함을 시사한다.

이를 구체화하고 실천하기 위해서는 사회과 외에도 다른 교과를 대상으로 연구를 확대하여, 교과별 의미 기능과 의미화 방식에 대한 체계화가 이루어져야 한다. 또한, 국어 의미학의 성과와의 연계, 그리고 실제 학습자의 어휘 발달 양상을 확인하는 실증적 연구도 필요하다. 본 연구를 비롯하여 후속 연구를 통해 국어과 어휘 교육이 학습자의 언어적 성장과 발달을 지원할 수 있기를 기대한다.

* 본 논문은 2025.01.26. 투고되었으며, 2025.02.11. 심사가 시작되어 2025.03.10. 심사가 종료되었음.

참고문헌

- 강형석·양장훈(2019), 「한국어 단어 임베딩을 위한 Word2vec 모델의 최적화」, 『한국디지털콘텐츠학회논문지』 20(4), 825-833.
- 김광해(2003), 『등급별 국어교육용 어휘』, 서울: 박이정.
- 김정인·박승규·박윤경·정요근·이종원·김자영·배성호·문재경·고아라·양은석·지태민·이은진·김민우·하늘빛·신혜선·김현애·신봉석(2022ㄱ), 『초등학교 사회 3-1』, 서울: 천재교육.
- 김정인·박승규·박윤경·정요근·이종원·김자영·배성호·문재경·고아라·양은석·지태민·이은진·김민우·하늘빛·신혜선·김현애·신봉석(2022ㄴ), 『초등학교 사회 4-1』, 서울: 천재교육.
- 김정인·박승규·박윤경·정요근·이종원·김자영·문재경·고아라·지태민·이은진·하늘빛·신혜선·신봉석(2023), 『초등학교 사회 6-1』, 서울: 천재교육.
- 박서빈·전영석·이창훈(2015), 「학교 학습에 의한 초등학생의 에너지 개념 변화 탐구」, 『에너지기후변화교육』 5(1), 1-7.
- 박세훈·장인실(2018), 「의미 연결망 분석을 통한 도덕과와 사회과의 다문화교육 개념 특성 비교 연구」, 『초등교육연구』 31(1), 47-72.
- 배희숙(2021), 『언어발달』, (2판), 서울: 학지사.
- 신명선(2004), 『국어 사고도구어 교육 연구』, 서울대학교 박사학위논문.
- 안혜지(2018), 「국어 다면어에 대한 재고찰」, 『반교어문연구』 49, 225-251.
- 윤길복(2012), 「사회과학 개념, 개념적 은유, 사회과 개념 학습」, 『사회과교육연구』 19(4), 45-62.
- 이민철(2024), 「Kiwi: 통계적 언어 모델과 Skip-Bigram을 이용한 한국어 형태소 분석기 구현」, 『디지털 인문학』 1(1), 109-136.
- 이수상(2012), 『네트워크 분석 방법론』, 서울: 논형.
- 이용준(2021), 「학문 문식성 교육의 필요성과 과제」, 『한국어문교육』 37, 43-88.
- 이윤미·김효정·송원섭·조철기(2024), 「언어 네트워크 분석을 활용한 통합사회 교과서의 행복 개념 분석」, 『한국지역지리학회지』 30(2), 307-319.
- 임지룡(2018), 『한국어 의미론』, 서울: 한국문화사.
- 임희석·고려대학교자연어처리연구실(2020), 『자연어처리 바이블: 핵심이론·응용시스템·딥러닝』, (수정판), 서울: 휴먼싸이언스.
- 장상호(1997), 『학문과 교육』, 서울: 서울대학교출판부.
- 전영주(2019), 「학습 문식성 신장을 위한 교과 어휘 교육 연구」, 이화여자대학교 박사학위논문.
- 전영주·김은성(2015), 「교과 학습 어휘의 개념화 및 기능 분류 연구」, 『새국어교육』 104, 45-72.
- 전영주(2020), 「내용교과를 위한 문식성에서의 어휘 교육 양상 고찰」, 『국어교육』 169, 195-233.

- 정한데로(2023), 「단어 사용의 변화와 다의어 생성」, 『언어와 정보 사회』 50, 213-250.
- 주세형(2021), 「국어교육의 새로운 내용론—학습의 본질 구현을 위한 초학문 융합 패러다임 설계」, 『국어교육』 174, 73-125.
- 정동권·연유상(2021), 「교과서와 상·하위 집단 학생 간의 태양 및 월식 의미망 분석 비교」, 『현장과학교육』 15(1), 35-58.
- 최성길·최원희·강창숙·박상준·최병천·조일현·권태덕·이수영·조성백·김상희·강봉균·정민정·김연주(2015), 『중학교 사회 2』, 서울: 비상교육.
- 최소영(2020), 「초등학교 교육용 어휘 등급화와 등급 보정 절차 연구」, 『국어교육학연구』 55(4), 223-254.
- 최소영·길호연·박태준·류수경(2022), 「교육용 텍스트의 이독성 질적 평가 준거 개발 및 타당화 연구」, 『국어교육』 176, 1-57.
- Daller, H., van Hout, R., & Treffers-Daller, J. (2003), "Lexical richness in spontaneous speech of bilinguals", *Applied Linguistics*, 24(2), 197-222.
- Fang, Z. (2012), "Language correlates of disciplinary literacy", *Topics in Language Disorders*, 32(1), 19-34.
- Fitzgerald, J., Elmore, J., Relyea, J. E., & Stenner, A. J. (2020), "Domain-specific academic vocabulary network development in elementary grades core disciplinary textbooks", *Journal of Educational Psychology*, 112(5), 855-879.
- Maaten, L. & Hinton, G. (2008), "Visualizing Data using t-SNE", *Journal of Machine Learning Research* 9(86), 2579-2605.
- Menczer, F., Fortunato, S., & Davis, C. A. (2022), 『네트워크 분석: 소셜 미디어에서 신경망 까지』, 손승우(역), 서울: 에이콘(원서출판 2016).
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013), "Efficient estimation of word representations in vector space", *International Conference on Learning Representations(ICLR)*, Scottsdale.
- Nippold, M. A. (2006), "Language Development in School-Age Children, Adolescents, and Adults", In Brown, K. (Ed.), *Encyclopedia of Language & Linguistics*, (2nd Ed.), Oxford: Elsevier Publishing.
- Sahlgren, M. (2008), "The distributional hypothesis", *Italian Journal of Disability Studies* 20, 33-53.
- Shanahan, T., & Shanahan, C. (2012), "What is disciplinary literacy and why does it matter?", *Topics in Language Disorders* 32(1), 7-18.
- Vygotsky, L.S. (2011), 『생각과 말: 심리학적 탐구』, 배희철·김용호(역), 서울: 살림터(원서출판 1934).

교과 어휘의 의미 구조 및 관계 변화 탐색 — 사회과 교과서 어휘의 워드 임베딩 및 네트워크 분석을 중심으로

전영주

본 연구는 사회과 어휘의 의미 구조 및 관계에 나타난 교과 어휘의 의미화 양상을 워드 임베딩과 네트워크 분석을 통해 탐색한 연구이다. 분석 결과 교과 어휘는 교과 지식 체계와 맥락 속에서 그 의미 기능과 의미화 방식이 구체화되고 있었으며, 학습의 과정은 교과 어휘가 구성하는 의미를 발견하고 재구성하는 과정이었다. 이를 통해 본 연구는 학습을 통한 어휘 발달은 곧 의미의 발달임을 확인하였으며, 국어과 어휘 교육 또한 결과 중심의 접근이 아닌 발달과 성장의 과정이라는 인식을 바탕으로 그 내용과 방법이 모색되어야 함을 강조하였다.

핵심어 어휘 교육, 워드 임베딩, 네트워크 분석, word2vec, 교과 문식성, 국어 의미 교육, 국어 교육

ABSTRACT

Exploring the Semantic Structure and Relations of Disciplinary Vocabulary

— Word Embedding and Network Analysis of Social Studies Textbooks

Chun Youngju

This study examined the semantic aspects of disciplinary vocabulary in social studies using word embedding and network analysis. The results indicate that disciplinary vocabulary evolves in function and meaning-making within subject knowledge system and context. Learning involves discovering and reconstructing meanings shaped by disciplinary vocabulary. This study confirmed that vocabulary development through learning is fundamentally the development of meaning, emphasizing that vocabulary education should be approached as a process of growth and development rather than as a result-centered approach.

KEYWORDS vocabulary education, word embedding, network analysis, word2vec, disciplinary literacy, Korean semantic education, Korean language education