



과학교과서 텍스트의 계량적 분석을 이용한 과학 개념어의 생산적 지식 교육 방안 탐색

윤은정*

경북대학교 과학교육연구소

Exploring Teaching Method for Productive Knowledge of Scientific Concept Words through Science Textbook Quantitative Analysis

Eunjeong Yun*

Science Education Research Institute of Kyungpook National University

ARTICLE INFO

Article history:

Received 13 February 2020

Received in revised form

18 February 2020

Accepted 24 February 2020

Keywords:

scientific concept, scientific concept word, word cluster, co-occurrence, text network analysis, word embedding

ABSTRACT

Looking at the understanding of scientific concepts from a linguistic perspective, it is very important for students to develop a deep and sophisticated understanding of words used in scientific concept as well as the ability to use them correctly. This study intends to provide the basis for productive knowledge education of scientific words by noting that the foundation of productive knowledge teaching on scientific words is not well established, and by exploring ways to teach the relationship among words that constitute scientific concept in a productive and effective manner. To this end, we extracted the relationship among the words that make up the scientific concept from the text of science textbook by using quantitative text analysis methods, second, qualitatively examined the meaning of the word relationship extracted as a result of each method, and third, we proposed a writing activity method to help improve the productive knowledge of scientific concept words. We analyzed the text of the "Force and motion" unit on first grade science textbook by using four methods of quantitative linguistic analysis: word cluster, co-occurrence, text network analysis, and word-embedding. As results, this study suggests four writing activities, completing sentence activity by using the result of word cluster analysis, filling the blanks activity by using the result of co-occurrence analysis, material-oriented writing activities by using the result of text network analysis, and finally we made a list of important words by using the result of word embedding.

1. 서론

모든 단어들은 형태와 개념을 갖는다. 그리고 아무리 간단한 단어라도 하나의 단어가 독립적으로 의미를 가질 수는 없으며 다른 단어 나 사물, 상황과의 연결 속에서 그 의미가 정해진다(Wittgenstein, 1958). 따라서 한 단어의 개념은 개념 요소를 이루는 여러 단어들이 모여서 특정 관계를 갖는 하나의 집합으로 볼 수 있다. 그리고 단어들이 형성하고 있는 이러한 네트워크는 단어 자체의 의미를 넘어서서 단어와 관련된 지식의 구조가 드러나는 표상으로 간주되기도 한다(Evans & Green, 2006). 같은 맥락에서 인지언어학자들은 언어의 구조, 즉, 단어들 사이의 관계들이 인간 인지의 직접적인 반영이라고 주장한다(Robinson, 2008). 따라서 언어 능력이 발달한다는 것은 동시에 인지 능력이 발달함을 의미하고, 이것은 곧 머리 속의 단어 네트워크가 양적, 질적으로 확장되고 잘 조직됨을 의미한다. 이러한 관점에서 과학 개념은 중심에 해당 개념을 표현하는 과학용어가 위치하는 하나의 단어 집합을 이루는 덩어리라고 볼 수 있다. 그리고 이러한 덩어리들 사이에 다시 복잡한 연결 관계를 맺으며 과학적 지식을 이루고 있다. 과학 교수학습에서는 하나의 개념 자체가 교육의 대상이

되는 경우도 있고, 개념들 사이의 관계가 대상이 되기도 한다.

위와 같은 관점에서 보면 과학 개념 학습에 있어 학생들의 머리 속에 해당 개념을 표현하는 과학용어와 그 용어를 둘러싼 단어 네트워크가 양적, 질적으로 잘 조직되는 것이 매우 중요하다. 이 때 양적, 질적으로 잘 조직된다는 것의 기준이 무엇인지에 대한 논의가 필요할 것이다. 아동의 언어발달, 혹은 개념 발달 과정을 탐색한 Piaget(1926) 나 Vygotsky(1962)의 연구에서는 정상의 성인들 사이에 통용되는 완성된 관념으로서의 단어의 의미를 개념이라고 부르고, 아동의 머리 속에 있는 성인의 것과는 완전히 일치하지 않는 관념들은 개념의 이전 단계(pseudoconcept)로 보고 있다. 같은 맥락에서 과학 개념을 살펴보면, 과학 개념의 완성도는 과학 분야의 전문가들 사이에 통용되는 관념으로 볼 수 있을 것이다. 그리고 전문가들의 완성된 개념은 그들이 사용하는 언어를 통해 표출된다. Yun & Park(2018)에서는 완성된 과학 개념을 구성하는 단어 네트워크를 '과학적 의미 네트워크(scientific semantic network)'로, 개별 학생들 각각이 가진 개념을 '머릿속 의미 네트워크(mental semantic network)'로 구분하고, 학교 과학교육에서 다루는 수준의 개념들의 과학적 의미 네트워크가 잘 반영되어 있는 자료로 과학교과서를 꼽은 바 있다. 요컨대 학생들이 과학 개념을 습득하기 위해서는 개념을 구성하고 있는 단어들 사이의

* 교신저자 : 윤은정 (ejyun@knu.ac.kr)
http://dx.doi.org/10.14697/jkase.2020.40.1.41

연결 관계를 습득하는 것이 중요하고, 과학교과서 텍스트에서 이상적인 단어 연결 모델을 찾을 수 있다. 본 연구에서는 과학 개념을 담은 단어를 과학 개념어라고 정의하고, 과학 개념어를 습득함에 있어 해당 개념의 과학적 의미 네트워크를 구성하는 단어 관계를 습득하는 것이 중요함을 배경으로 하고 있다.

한편, 하나의 단어를 안다는 것은 단어의 형태와 의미, 용법에 관한 지식, 정확하고 적절하게 사용하는 능력 등을 포함한다(Lee, 1997). 그리고 단어의 학습 측면에서 이들은 단어에 대한 수용적 지식과 생산적 지식으로 구분하기도 한다(Kim & Im, 2012). 단어에 대한 수용적 지식은 단어의 의미를 이해하고 기억 속에 저장하거나 회상하는 능력으로 읽기나 듣기와 관련이 있고, 생산적 지식은 단어를 적절한 시기와 상황에 정확하게 사용할 수 있는 능력으로 쓰기, 말하기와 관련이 있다(Nattinger, 1988; Nation, 1990; Laufer, 1998). 단어에 대한 수용적 지식은 시간적 효율성의 측면에서 단기간에 많은 단어를 습득할 수 있다는 점(Folse, 2006)과 단어에 대한 양질의 정보가 입력된다는 점(Shintani, 2011)에서 중요하고, 단어에 대한 생산적 지식은 실제 의사소통에 필수적인 기능이라는 점과 단어에 대한 깊은 이해와 정확한 처리 능력을 가져온다는 점(Hulstijn & Laufer, 2001)에서 각기 중요한 의미를 가진다. 그런데, 단어에 대한 수용적 지식과 생산적 지식은 상호의존적이기 보다는 독립적인 기능으로, 한 쪽이 다른 쪽을 보장해 주지 않는다(Griffin & Harely, 1996; Schmitt, 2000; Waring, 1998). 즉, 수용적으로 습득된 단어는 수용적 지식에 머무르며, 생산적 지식으로 발전하기 위해서는 쓰기 등의 활동을 통해 생산적으로 단어를 사용할 수 있는 기회를 주는 등 별도의 학습이 필요하다(Ryoo, 2009; Zhou, 2010). 특히 생산적 단어 학습 가운데 쓰기 활동은 학생들의 머리 속에 단어의 의미를 정교화하는 데 중요한 역할을 하는 것으로 알려져 있다. 단어를 문장에 사용하기 위해서는 그 단어와, 단어가 사용되었던 문맥을 충분히 회상해야 하고, 문장에서 적절하게 단어를 배치하기 위해서는 의미의 적절성을 평가하는 과정을 거쳐야 하기 때문이다(Barcroft, 2004).

본 연구에서는 과학 개념어에 대한 교육을 언어학적 관점에서 바라보고, 과학 개념을 구성하고 있는 단어들 사이의 관계를 생산적이고 효과적으로 교육할 수 있는 방안을 탐색함으로써, 학생들이 과학 개념어에 대한 깊고 정교한 이해와 더불어 정확하게 사용할 수 있는 능력을 길러주는 것을 궁극적인 목적으로 하고 있다. 현재 사용되고 있는 과학 교과서들에도 일부 과학 개념어에 대한 생산적 지식 교육이 이루어지고 있기는 하다. 2015 개정 교육과정에 따른 중학교 교과서들 가운데는 특정 개념어들을 제시하고, 이들이 포함된 문장을 작성해보도록 하는 활동을 제시하는 경우를 찾아볼 수 있다. 그러나 해당 개념어에 대한 기초적인 생산적 지식 교육 없이 이러한 열린 활동을 제시할 경우 학생들이 작성한 문장의 정확도나 질이 낮을 수밖에 없고, 교사가 모든 학생의 글을 반복적으로 교정하고 평가하기 어려우므로 교육적 효과를 크게 기대하기는 어려울 것이다. 또한 단 한 번의 문장 만들기 활동으로 과학 개념어에 대한 깊이 있는 이해가 이루어지는 것은 무리이므로 보다 효과적이고 체계적인 방안을 모색할 필요가 있다. 현장의 과학 교사들 또한 과학 개념어에 대한 교육의 중요성에 충분히 공감하고 있음에도 불구하고 구체적인 교육 방법에 대한 정보가 없어서 어려움을 겪고 있음을 호소한 바 있다(Yun *et al.*, 2015).

이에 본 연구에서는 첫째, 몇 가지의 계량 언어학적 텍스트 분석 방법을 이용하여 과학 교과서 텍스트로부터 과학 개념을 구성하고 있는 단어들과 그들 사이의 관계를 추출하고, 둘째, 각 방법의 결과로 추출된 단어 관계의 의미를 정성적으로 살펴본 뒤, 셋째, 이를 이용하여 과학 개념어의 생산적 지식 향상에 도움을 줄 수 있는 쓰기 활동 방법을 제안해 보고자 하였다.

II. 연구 방법

1. 분석 대상

과학 개념을 구성하고 있는 단어들과 이들 사이의 관계를 알아보기 위한 대상 개념어로 물리 분야의 개념어인 ‘힘’을 선정하였다. ‘힘’은 물리학에서 기본적인데도 중요한 개념을 담고 있는 개념어이면서, 일상적 의미와의 혼용으로 인해 학생들이 정확한 개념 습득에 어려움을 겪는 경우가 많은 개념어 가운데 하나이다(Rincke, 2010; Yun & Park, 2016). ‘힘’이 사용된 과학교과서 텍스트로는 6차, 7차, 2009 개정 과학교육과정에 따른 과학교과서 총 24권으로 부터 ‘힘과 운동’ 단원의 텍스트를 추출하여 사용하였다. 텍스트를 계량 언어학적 방법으로 분석하기 위해서는 전처리 과정이 필요한데, 먼저, 형태소 분석기 Utagger(Shin & Ock, 2012)를 이용하여 형태소 분석을 실시하고, 조사와 어미류를 삭제한 뒤, 동사나 형용사 등의 용언들은 모두 기본형으로 변환하였다. 최대한 자연스러운 상태의 단어 관계를 살펴보고자 하였으므로 그 외의 문장을 구성하는 단어들은 가능한 한 그대로 보존하고자 하였다.

2. 텍스트 분석 방법

본 연구에서는 과학 교과서의 텍스트 분석 방법으로 단어 클러스터 분석, 공기 관계 분석, 텍스트 네트워크 분석, 워드임베딩의 네 가지를 선택하였다. 이들 분석 방법은 언어학 분야에서는 광범위하게 사용되고 있는 방법들로서 방법 자체가 갖는 의미보다는 텍스트의 특성에 따라 나타나는 분석 결과의 특이성에 초점을 맞추어 활용 방안을 찾는 것이 중요하다.

가. 단어 클러스터 분석

단어 클러스터 분석은 텍스트 내에서 관련된 단어들의 집단을 추출한다는 포괄적인 의미를 가지고 있으며, 실제로 무엇을 ‘관련 단어’로 볼 것인가, 혹은 어느 범위까지를 ‘집단’으로 볼 것인가에 따라 다양하게 사용되고 있다(Jhang, S., & Lee, S. 2013; Lee & Kim, 2006). 본 연구에서는 출현 빈도와 분포를 기반으로 추출한 ‘단어 묶음’의 개념으로(Biber *et al.*, 1999) 클러스터의 의미를 제한하여 사용하였다. 즉 분석 대상 텍스트에서 한 문장 내에 과학 개념어인 ‘힘’과 좌우의 지정된 범위 내에서 함께 사용된 어절을 묶은 다음 텍스트 전체에서 나타난 묶음의 빈도를 기준으로 분석한다. 따라서 단어 클러스터 분석 결과에서 고빈도로 사용되는 클러스터는 해당 단어와 함께 사용되는 매우 전형적인 표현으로 해석할 수 있다. 일반적으로 언어학에서는 숙어와 같은 정형적 표현을 탐색하는 데 활용하기도 한다. 본

연구에서는 AntConc 프로그램을 사용하여 어절 단위의 단어 클러스터를 분석하였다.

나. 공기 관계 분석

공기(共起)는 두 단어가 같은 문맥에서 함께 사용되는 현상을 의미하며, 공기의 범위는 연구 목적에 따라 앞뒤 한 어절에서부터 한 문단에 이르기까지 다양하게 정할 수 있다(Park, 2002). 두 단어가 공기 관계에 있다고 함은 한 단어가 다른 한 단어와 함께 나타나도록 기대되는 빈도에 비해 실제 함께 나타나는 빈도가 높음을 의미한다. 일반적으로 언어 교육을 위해 단어의 사용 형태적, 의미적으로 연관 단어를 탐색하는 방법으로 매우 널리 사용되는 분석 방법이다. 공기 관계는 단순히 공기 빈도로 이야기하기도 하고, 공기 관계 지수로 나타내기도 하는데 공기 관계를 나타내는 지수로는 T-점수, Z-점수, MI, LMI, 카이제곱 등이 사용된다(Kang, 2011). 본 연구에서는 과학 개념어 ‘힘’과 공기 관계에 있는 단어들을 추출함에 있어 공기 빈도와 함께 MI 값을 사용하였다. MI 가 다른 값들에 비해 저빈도 공기어에 대해 비교적 높게 평가하는 경향이 있어서(Barnbrook, 1996), 정형적인 표현이지만 저빈도로 사용된 경우를 잘 드러내는 것으로 알려져 있다. MI 값을 구하는 방법은 아래와 같다(Stubb, 1995).

$$MI = \log_2 \{ [f(n,c) \times N] / f(n)f(c) \}$$

위의 식에서 $f(n)$ 은 전체 텍스트에서 대상 단어가 나타난 빈도, $f(c)$ 는 전체 텍스트에서 공기어가 나타난 빈도, $f(n,c)$ 는 전체 텍스트에서 두 단어가 함께 나타난 빈도, N 은 전체 텍스트의 어절수를 의미한다. 공기 관계 역시 AntConc 프로그램을 사용하였다.

다. 텍스트 네트워크 분석

텍스트 네트워크 분석 방법은 복잡계 이론을 언어에 적용하여 단어 사이의 관계를 살펴보는 방법으로서(Shin *et al.*, 2010) 단어들 사이의 관계 및 네트워크를 시각적으로 들여다볼 수 있다는 점, 텍스트 내에서 개별 단어들의 영향력 혹은 중요성을 정량적으로 살펴볼 수 있다는 점 등을 특징으로 꼽을 수 있다. 본 연구에서는 과학 개념어 교육을 위해서 과학 개념을 구성하고 있는 단어들 및 단어 사이의 관계, 네트워크적 특징 등을 살펴보는 것에 목적이 있으므로 네트워크의 시각화는 사용하지 않고 네트워크 지표들을 중심으로 분석해 보았다. 텍스트 네트워크를 구성하기 위한 조건으로는 문장 내에서 한 단어를 기준으로 좌우 5어절 내에 있는 단어들을 연결 관계가 있는 것으로 보고, 이 때 어절 거리가 가까울수록 강한 관계가 있는 것으로 간주하여 어절 거리에 가중치를 부여하였다. 네트워크 분석은 Gephi 0.9.2를 사용하였고 네트워크 분석 지표는 연결중심성, 근접중심성, 사이중심성, 고유벡터중심성의 네 가지를 살펴보았다. 연결중심성은 하나의 단어가 다른 단어와 얼마나 많이 연결되어 있는가를 나타내는 지표이고, 근접중심성은 네트워크에서 위상적으로 얼마나 중심에 있는가를 나타내는 지표이다. 그리고 사이중심성은 한 단어가 다른 단어들을 매개해주는 정도를 의미하는 지표이고, 마지막으로 고유벡터중심성은 한 단어가 얼마나 중요한 단어와 연결되어 있는가를 나타내는 지표로 해석할 수 있다(Wasserman & Faust, 1994).

라. 워드임베딩

워드임베딩은 언어학의 분산가설을 전제로(Sahlgren, 2008) 특정 어휘에 속하는 모든 단어를 벡터 공간상의 실수 벡터로 매핑하는 기법으로서, 분석 결과가 단어의 문법 정보와 의미 정보를 매우 잘 나타내고 있어(Choi, 2018), 언어학 및 자연언어처리 분야 등에서 많은 관심을 받고 있다. 워드임베딩의 결과물은 분야에 따라 다양한 측면에서 활용되는데, 본 연구에서는 연구 대상 텍스트를 이용하여 워드임베딩을 실시했을 때 과학 개념어 ‘힘’과 유사한 벡터값을 갖는 단어 목록을 추출하는 데 활용하였다. 그리고 추출된 단어들과 ‘힘’과의 관련성을 살펴 ‘힘’과 관계된 단어 목록을 작성해 보았다. 워드임베딩 모델로는 Word2vec(Mikolove *et al.*, 2013)을 사용하였고, 도구로는 Python 3.6의 Gensim 라이브러리를 사용하여 수행하였다.

III. 연구 결과 및 논의

1. 어절 단위 클러스터 분석

가. 분석 결과

어절 단위의 클러스터 분석을 통해 과학용어 ‘힘’을 기준으로 좌우 각각 2어절에서 4어절까지의 어절 연쇄 추출해 본 결과 표 1, 표 2와 같이 나타났다. 표 1은 ‘힘’을 기준으로 오른쪽으로 연이어 자주 함께 사용되는 어절 연쇄들을 제시하였고 표 2에는 ‘힘’을 기준으로 왼쪽으로 연이어 자주 함께 사용되는 어절 연쇄들을 제시하였다. 표 1과 2에 추출된 표현들은 일상어에서는 잘 나타나지 않는 형태로(Yun & Park, 2016), 과학적 의미로서의 ‘힘’이 사용될 때 보이는 매우 정형적인 표현들에 해당한다. 예를 들어 2어절에서 가장 높은 빈도를 차지한 ‘힘 작용하다’의 연쇄는 과학적 의미에서의 ‘힘’을 기술하는데 가장 중요하고 기본적인 주술관계라고 할 수 있다. 일상적 의미에서는 주로 ‘힘 세다’, ‘힘 주다’의 표현이 많이 사용되는데(Yun & Park, 2016), 학생들의 머리 속에서 ‘힘’의 일상적 의미와 과학적 의미를 잘 구분하고 과학적 맥락에 친숙해지도록 하기 위해서는 이러한 기본적인 주술관계에 익숙해지도록 하고 이후 확장해 나가는 과정이 중요하다. 따라서 어절 단위 클러스터 분석 결과는 ‘힘’을 처음 배우기 시작하는 단계에서 ‘힘’의 과학적 의미를 일상적 의미와 분리시키고, 과학적 의미를 활성화 시키는 토대를 마련하기에 좋은 자료가 된다.

나. 클러스터 분석 결과를 이용한 문장 완성하기 활동

클러스터 분석에서는 개념 학습 초기 단계에 활용할 수 있는 어절 연쇄를 이용한 문장 완성하기 기법을 소개하고자 한다. 표 1과 2의 어절 연쇄를 구성하고 있는 어휘들을 살펴보면 일정 어휘들이 반복적으로 사용되면서 다양한 연쇄 구성을 이루고 있음을 알 수 있다. 그리고 이러한 어절 연쇄들은 과학교사나 과학교육 전문가들이 어느 정도 연역적 예측이 가능할 수도 있었으나, 사용 빈도 정보를 고려한 목록은 클러스터 분석과 같은 정량적 방법을 이용해서만이 추출이 가능하다.

‘힘’을 둘러싼 어절 연쇄를 구성하고 있는 어휘들을 살펴보면 ‘작

Table 1. Word cluster from the 2nd to 4th right of 'force'

순위	2어절	빈도	3어절	빈도	4어절	빈도
1	힘 작용하다	1022	힘 작용하다 때	196	힘 작용하다 때 물체	57
2	힘 합력	432	힘 작용하다 물체	172	힘 작용하다 물체 속력	49
3	힘 받다	307	힘 평형 이루다	88	힘 합력 어떻게 구하다	47
4	힘 크기	286	힘 합력 구하다	87	힘 반대 방향 작용하다	40
5	힘 방향	250	힘 받다 물체	69	힘 작용하다 물체 운동	37
6	힘 주다	220	힘 방향 같다	57	힘 동시 작용하다 때	32
7	힘 물체	159	힘 합력 어떻게	54	힘 받다 물체 운동	32
8	힘 운동	136	힘 작용하다 때문	50	힘 같다 방향 작용하다	30
9	힘 평형	110	힘 반대 방향	49	힘 합력 구하다 수	28
10	힘 어떻게	104	힘 작용하다 것	49	힘 작용하다 때 힘	25

Table 2. Word cluster from the 2nd to 4th left of 'force'

순위	2어절	빈도	3어절	빈도	4어절	빈도
1	작용하다 힘	798	물체 작용하다 힘	192	나란하다 방향 작용하다 힘	48
2	물체 힘	480	방향 작용하다 힘	145	물체 운동 방해하다 힘	44
3	방향 힘	307	수 있다 힘	101	반대 방향 작용하다 힘	43
4	때 힘	190	같다 크기 힘	81	같다 방향 작용하다 힘	41
5	뉴턴 힘	174	사이 작용하다 힘	77	물체 사이 작용하다 힘	37
6	크다 힘	156	운동 방해하다 힘	75	크다 힘 작다 힘	27
7	크기 힘	153	반대 방향 힘	66	힘 작용하다 때 힘	25
8	있다 힘	149	작용하다 때 힘	63	알다 수 있다 힘	23
9	당기다 힘	134	나란하다 작용하다 힘	58	운동방향 반대 방향 힘	22
10	나란하다 힘	107	같다 방향 힘	52	운동방향 나란하다 방향 힘	22

용하다', '크기', '방향', '운동', '같다' 등의 어휘처럼 좌우 어절 연쇄에 공통적으로 사용되는 경우도 있고, '합력', '평형', '이루다', '구하다' 처럼 '힘'의 왼쪽에만 위치하는 경우와 '당기다', '나란하다', '방해하다', '크다' 와 같이 '힘'의 오른쪽에만 위치하는 어휘들도 있었다. 표 1과 2에 제시된 어절 연쇄를 이용한 문장 완성하기의 예시를 아래에 제시하였다. 학생들에게 어절 연쇄를 제시하고 문장이나 구를 만들도록 할 수 있다. 경우에 따라서는 왼쪽의 어절 연쇄와 오른쪽 어절 연쇄를 조합하여 제시할 수도 있을 것이다.

ㄱ. 힘 + 작용하다

- 힘이 작용하다
- 물체에 힘이 작용하다
- 고무공에 힘을 작용하다

ㄴ. 작용하다 + 힘

- 고무줄에 작용하는 힘이 크면 많이 늘어난다
- 물체에 작용하는 힘이 평형을 이룬다.
- 용수철에 작용하는 힘과 늘어난 길이 사이에 비례 관계가 있다.

ㄷ. 힘 + 평형 + 이루다

- 책상위의 컵에 작용하는 중력과 책상이 받치는 힘이 평형을 이루다
- 정지한 물체에 작용하는 힘들은 평형을 이루고 있다

ㄹ. 운동 + 방해하다 + 힘

- 운동을 방해하는 힘
- 자전거의 운동을 방해하는 힘에는 마찰력이 있다.
- 낙하하는 공의 운동을 방해하는 힘에는 공기저항이 있다.

ㅁ. 힘 + 작용하다 + 물체 + 속력

- 힘이 작용하면 물체의 속력이 변한다
- 운동하는 물체에 반대 방향으로 힘이 작용하면 물체의 속력이 줄어든다.

ㅂ. 나란하다 + 방향 + 작용하다 + 힘

- 운동방향과 나란한 방향으로 작용하는 힘
- 운동방향과 나란한 방향으로 작용하는 힘 중에는 같은 방향으로 작용하는 힘과 반대 방향으로 작용하는 힘이 있다.

ㄱ과 ㄴ은 '힘'과 '작용하다'의 두 어휘가 순서만 다른 경우의 예시이다. ㄱ의 경우는 어절 연쇄의 앞으로 확장 가능성이 높고 ㄴ의 경우는 어절 연쇄의 뒤로 확장 가능성이 높다. ㄷ에서 ㅂ까지의 예시를 살펴보면 어절 연쇄를 구성하는 어절 수가 증가할수록 '힘'과 관련된 보다 구체적인 물리적 상황으로 제한됨을 알 수 있다. 학생들이 제시된 어절 연쇄가 포함된 문장을 완성하기 위해서는 적절한 물리적 상황을 스스로 탐색해야 하고, 물리적 현상이 포함된 다양한 실제 사례들을 떠올려야 한다. 따라서 이 과정에서 물리적 현상에 대한 이해와 실제 상황에서의 적용, 그리고 '힘'에 대한 개념 이해가 종합적으로 어

우리질 수 있다. 특정한 물리적 상황을 다루는 수업에서는 상황에 적합한 어절 연쇄를 선택하여 제시할 수 있고, 힘과 관련된 물리적 상황을 보다 발산적으로 떠올리기를 유도하려면 기어나 ㄴ과 같은 2어절 연쇄를 제시하는 것이 효과적일 수 있을 것이다.

2. 공기 빈도 분석

가. 분석 결과

공기는 문장 내에서 특정 단어들 이 함께 사용되는 경우를 의미하는 데, 이 때 함께 사용되는 범위나 거리는 연구의 목적에 따라 적절하게 정할 수 있다. 예를 들어 특정 단어를 기준으로 좌우 3어절 내에 함께 사용되는 경우로 정할 수도 있고, 좌우 5어절 이내를 범위로 정할 수도 있다. 또, 특정 단어로 부터 한 어절 건너, 혹은 두 어절 건너 자리에 함께 사용되는 경우로 정하거나, 방향까지 지정하여 특정 단 어 왼쪽 세 번째 어절 혹은 오른쪽 두 번째 어절 등으로 정할 수도 있다. 아래 표 3은 ‘힘’을 기준으로 좌우 각각 첫 번째 어절에서 부터 세 번째 어절까지의 자리에 많이 사용되는 단어들을 상위 20개씩 빈 도순으로 제시한 것이다.

나. 공기 빈도 분석 결과를 이용한 빈칸 채우기 활동

여기서는 공기 빈도 분석 결과를 이용하여 빈칸 채우기 형태의 문장 만들기 기법을 제시해보고자 한다. 표 3에서 좌우 각각 첫 번째 어절의 분석 결과는 표 1과 2의 2어절 연쇄와 결과 값이 같으므로 중복된 언급을 피하기 위하여 다루지 않았다. 한편, 같은 단어라고 하더라도 위치하는 자리에 따라서 전혀 다른 문장이 만들어질 수 있 고 머리 속에 유도되는 물리적 상황이 달라지게 된다. 표 3에서 보면

‘물체’는 ‘힘’의 오른쪽 첫 번째 어절을 제외한 나머지 위치에 매우 높은 빈도로 사용되고 있다. ‘물체’가 ‘힘’의 오른쪽 첫 번째 어절에서 는 상대적으로 잘 사용되지 않으며, 따라서 ‘힘 + 물체’의 어절 연쇄를 제시하면 상대적으로 문장의 활용도가 낮아지므로 다른 자리에 비하 여 답이 제한적일 수 있다. 아래에 ‘물체’가 ‘힘’의 좌우 각각 두 번째 어절과 세 번째 어절에 제시되었을 때의 문항과 각각의 답안 예시를 제시하였다.

ㄸ. 힘 + () + 물체

힘을 (받는) 물체는 모양이 변하거나 운동상태가 변한다.
오른쪽으로 운동하는 물체에 오른쪽으로 3N의 힘을 (가하면)
물체의 속력이 더 커진다.

ㅇ. 힘 + () + () + 물체

이것은 힘을 (받아서) (변형된) 물체이다.
물체의 운동방향과 힘의 (방향)이 (반대이면) 물체의 속력은 줄어든다.

ㅈ. 물체 + () + 힘

두 물체 (사이에) 힘이 작용하다.
지구가 물체를 (당기는) 힘을 중력이라고 한다.

ㅊ. 물체 + () + () + 힘

물체의 (운동을) (방해하는) 힘에는 마찰력이 있다.
두 물체 (사이에)는 (끌어당기는) 힘이 작용한다.

위의 예시는 같은 단어가 위치하는 자리에 따라서 다양한 문장이 만들어질 수 있으며, 이러한 변화를 통해 다양한 학생들의 사고를

Table 3. Highly collocated words with ‘force’ from 1st to 3rd right and left

1R	빈도	2R	빈도	3R	빈도	1L	빈도	2L	빈도	3L	빈도
작용하다	1022	물체	454	물체	333	작용하다	798	물체	409	물체	339
합력	432	때	314	작용하다	205	물체	480	방향	321	방향	205
받다	307	방향	237	운동	177	방향	308	작용하다	212	같다	170
크기	286	하다	223	방향	170	때	190	같다	178	작용하다	137
방향	250	작용하다	198	때	142	뉴턴	174	운동	138	있다	135
주다	220	것	153	속력	130	크다	156	사이	135	하다	124
물체	159	같다	142	수	127	크기	155	있다	124	때	120
운동	136	속력	120	하다	125	있다	149	때	122	운동	105
평형	110	어떻다	116	있다	114	당기다	134	합력	114	합력	95
어떻다	104	이루다	98	나타내다	97	나란하다	107	나란하다	113	나타내다	79
같다	96	나타내다	94	같다	91	가지	96	수	110	운동방향	77
가하다	92	운동	93	화살표	71	같다	89	것	85	수	77
화살표	91	구하다	90	변하다	71	방해하다	88	화살표	78	것	75
나타내다	85	때문	81	그림	69	하다	78	하다	78	되다	73
하다	82	화살표	78	것	69	밀다	75	용수철	77	나란하다	70
크다	81	크기	74	합력	67	끌어당기다	74	크기	72	크기	62
효과	75	관계	68	구하다	66	나타내다	71	반대	69	반대	61
동시	53	있다	57	되다	61	것	68	되다	69	변하다	53
들다	51	수	55	모양	60	운동	62	운동방향	55	지구	51
반대	49	되다	55	움직이다	56	되다	60	효과	52	어떻다	51

자극할 수 있을 것으로 기대할 수 있다. 이 때 어떤 단어를 선택할 것인가는 수업 목표에 따라서 교사가 선택할 수 있으며, 표 3과 같은 위치별 공기어 목록이 활용될 수 있다. 사용 빈도가 높은 단어와 자리를 제시할수록 학생들은 문장을 구성하기에 용이할 것이고, 또 다양한 답이 나올 수 있다. 반대로 빈도가 낮은 단어와 자리를 제시하면 문장의 활용도가 낮으므로 학생들의 답은 보다 구체적이고 제한적인 상황으로 유도될 수 있다. 따라서 교사의 의도에 따라 적절하게 활용될 수 있다.

한편, 표 4에는 ‘힘’과의 거리를 고려하지 않은 공기어들을 나타내 보았다. 앞에서 제시한 단어와 위치를 함께 지정해주는 활동이 물리적 상황을 다소 제한적으로 유도할 수 있다면, 표 4의 경우는 어느 정도 ‘힘’과 관련된 물리적 상황에 익숙해진 학생들에게 보다 자유롭고 발산적으로 문장을 구성해볼 수 있도록 하기에 적절하다고 여겨진다. 이 때 ‘힘’을 기준으로 왼쪽이나 오른쪽이나의 방향에 관계 없이 제시할 때는 표 4의 전체 공기어 빈도 목록과 값을 참고할 수 있고, 방향을 지정해 주고 싶다면 왼쪽 공기어 빈도 또는 오른쪽 공기어 빈도를 선택적으로 활용할 수 있다. 한편, 공기지수 MI는 단어의 출현 기대 빈도 대비 관측 빈도를 고려한 값으로서 텍스트 전체에서 상대적으로 ‘힘’ 주위에만 특징적으로 모여 있는 단어들을 의미한다. 사용 빈도는 높지 않더라도 텍스트의 다른 부분에는 잘 나타나지 않고 ‘힘’과 같은 문장 내에서만 특징적으로 사용될 경우 공기지수는 높게 나타난다. 공기지수가 높은 단어들은 ‘힘’과 함께 제시하여 문장을 구성하도록 할 수도 있고, 앞의 7에서 7까지와 같은 과제를 제시할 때 참조 단어로 제시할 수도 있을 것이다.

3. 네트워크 분석

가. 분석 결과

네트워크 분석은 힘과 운동 단위 텍스트 전체에서 ‘힘’과 다른 단어들이 어떻게 유기적으로 연결되어 있는지에 대한 정보를 제공한다. 앞의 클러스터와 공기어 분석 결과들이 ‘힘’을 문장 단위에서 다양하게 사용하는 연습에 유용한 정보를 제공했다면 네트워크 분석 결과들은 문단 이상의 보다 넓은 범위에서 ‘힘’을 서술할 때 중요하게 사용되는 단어들의 정보를 제공한다. 통상적으로 텍스트 네트워크 분석 후에는 결과를 시각적으로 나타내어 단어들 사이의 복잡한 연결 구조를 보여주는 경우가 많으나, 본 연구에서는 힘과 운동 단위 텍스트의 네트워크 구조를 살피기보다는 결과의 교육적 활용에 초점을 두고 있으므로 시각적 결과물 대신 네트워크 지표들을 중심으로 논의하고자 한다. 아래 표 5는 힘과 운동 단위 전체의 텍스트 네트워크 가운데 ‘힘’을 중심으로 한 하위 네트워크에서 네트워크 지표별 상위 20개 단어들을 제시한 것이다. 연결중심성을 제외한 나머지 근접중심성, 사이중심성, 고유벡터중심성의 결과들은 앞서 클러스터 분석이나 공기 빈도 분석의 결과에서는 드러나지 않았던 단어들을 알 수 있다.

표 5에서 각 지표별 결과에 대해 해석해보면 다음과 같다. 먼저 연결중심성은 얼마나 많은 단어들과 많이 연결되어 있는가를 나타내는 지표로서 텍스트 내의 빈도 정보를 가장 많이 반영하는 지표라고 볼 수 있다. 즉, 표 5에서 연결중심성이 높은 ‘물체’, ‘속력’, ‘운동’, ‘작용하다’ 등의 단어들은 ‘힘’을 설명할 때 가장 많이 사용하는 단어들에 해당한다. 근접중심성은 네트워크의 위상적(topology) 측면에서 얼마나 중심에 있는가를 나타내는 지표로서 주변 단어들 사이에서 중심에 위치한 정도를 의미한다. 네트워크를 시각화 한다면 전체 네

Table 4. Words cooccurrence with ‘force’ excluded from distance

전체 공기어	빈도	왼쪽 공기어	빈도	오른쪽 공기어	빈도	공기지수 상위어	MI
작용하다	3378	물체	1800	작용하다	1812	효과	5.63
물체	3250	작용하다	1566	물체	1450	합하다	5.53
방향	2158	방향	1158	방향	1000	작용점	5.48
때	1345	때	683	합력	801	요소	5.47
있다	1242	있다	666	운동	670	합성	5.38
운동	1227	같다	614	하다	664	평형	5.36
합력	1213	운동	557	때	662	합성하다	5.28
하다	1151	하다	487	있다	576	의미	5.27
같다	1142	크기	451	크기	572	합력	5.21
크기	1023	것	431	같다	528	더하다	5.14
것	847	합력	412	어떻다	440	화살표	5.07
수	751	수	352	속력	426	주다	5.04
나타내다	718	나란하다	336	것	416	가하다	5.00
어떻다	698	되다	333	수	399	작용	4.95
속력	611	나타내다	328	나타내다	390	빠다	4.93
되다	587	뉴턴	323	받다	389	나란하다	4.93
화살표	550	사이	297	화살표	345	이상	4.90
그림	542	당기다	286	그림	310	변	4.86
받다	525	크다	264	주다	293	이루다	4.84
뉴턴	516	변하다	259	변하다	255	작용하다	4.81

트위크를 구성하고 있는 작은 집단들이 있고, 각각의 작은 집단들의 중심에 위치한 단어가 근접중심성이 높은 단어에 해당한다. 텍스트의 내용을 고려한다면 근접중심성이 높은 단어들은 텍스트를 구성하고 있는 소주제들의 중심어의 역할인 셈이다. 따라서 표 5에 제시된 근접중심성 상위 단어들은 ‘힘’을 설명하는 과정에서 동원된 여러 가지 작은 스토리들의 중심 소재로 볼 수 있을 것이다. 근접중심성은 최소 0에서 최대 1까지의 값을 갖는데 표 6에 제시된 20개의 단어는 모두 근접중심성 최대값 1을 가지고 있으며 이들 가운데 연결중심성이 높은 단어 순서로 정렬한 것이다. 전체 분석 결과에서는 전체 수천 개의 단어들 가운데 근접중심성 1을 갖는 단어는 약 40여개가 나타났는데, 이는 힘과 운동 단원이 ‘힘’을 중심으로 약 40여개의 작은 소주제로 구성되어 있음을 의미한다고 볼 수 있다. ‘주기’, ‘알짜힘’, ‘진폭’ 등은 ‘힘’과 관련된 하나의 독립된 개념으로서의 이야기 소재로 볼 수 있을 것이다. ‘주기’와 ‘진폭’의 경우 맥락상 ‘진자’와 관련된 개념들이므로 ‘진자’가 중심 소재로 나타날 것으로 기대할 수 있으나, 실제로 ‘힘’ 단원에서는 진자 자체보다는 진자의 ‘주기’와 ‘진폭’에 영향을 주는 요인에 초점을 맞추어 기술하는 경향으로 인해 ‘진자’ 대신 ‘주기’와 ‘진폭’이 이야기의 중심으로 추출된 것으로 해석할 수 있다. 또, ‘탄성체’, ‘말’, ‘로켓’, ‘스케이트’ 등은 힘의 작용이 잘 드러나는 물리적 상황으로서의 이야기 소재로 볼 수 있을 것이다. ‘동영상’의 경우 동영상이라는 단어 자체보다는 동영상에 담긴 내용이 이야기 소재로 사용되었을 것을 추측할 수 있다. 실제로 교과서 텍스트를 살펴본 결과 주로 낙하하는 물체의 동영상, 운동하는 물체의 동영상을 이용하여 물체의 운동을 분석하고 힘과 관련지어 이야기하는 부분이 많았다.

다음으로 사이중심성이 높은 단어들은 네트워크에서 단어들 사이의 연결을 매개하는 역할을 하는 경우이다. 즉, 특정 물리적 상황을

‘힘’과 관련지어 설명할 때, 상황과 ‘힘’과의 연결고리로 볼 수 있을 것이다. 예를 들어 근접중심성이 높은 ‘말’, ‘로켓’, ‘스케이트’를 소재로 ‘힘’에 대해서 이야기하기 위해서는 이들의 ‘운동’을 분석하여 이야기해야 할 것이고 ‘속력’의 변화도 이야기해야 할 것이다. 또, 지면으로 모든 상황을 설명해야 하는 과학교과서의 특성상 물리적 상황에 대한 그림을 이용하여 설명하는 경우가 많았는데 이 때문에 ‘그림’이 사이중심성이 높게 나타난 것으로 해석할 수 있다. 끝으로 고유벡터 중심성은 얼마나 영향력 있는 단어들과 연결되어 있는가를 나타내는 지표인데, 영향력이 높다는 것은 주로 연결중심성이 높음을 의미한다. 따라서 고유벡터 중심성이 높은 단어들을 사용하면 연결중심성이 높은 단어와 연결하기 쉽다는 것을 의미한다. 연결중심성이 높은 단어들은 사용 빈도가 높은 단어들이므로 고유벡터 중심성이 높은 단어를 사용하면 그만큼 문장을 구성하기 용이하게 된다.

나. 네트워크 분석을 이용한 소재 중심 글쓰기 활동

네트워크 분석 결과들을 이용하여 활용할 수 있는 방안으로 소재 중심 글쓰기 기법을 생각해볼 수 있다. ‘힘’과 관련된 수업이므로 주제는 ‘힘’이 될 것이고, 근접중심성 지표값을 참고하여 글의 소재를 선택할 수 있다. 그리고 사이중심성이 높은 단어들 중에서 상황에 적절한 단어들을 선택하여 필수 단어로 제시해 주면 ‘힘’과 관련된 다른 단어들을 유도하는 효과를 기대할 수 있다. 다음으로 고유벡터 중심성이 높은 단어를 먼저 선택하도록 하고, 이어 선택한 단어와 연결할 단어를 연결중심성 상위 단어 가운데 선택하도록 한다. 이렇게 소재 선택, 필수 단어 선택, 문장 구성을 위한 참조 단어쌍 선택의 순서로 제시하여 학생들로 하여금 ‘힘’과 관련된 물리적 상황을 기술

Table 5. Words with high centrality in the text of ‘force’

연결중심성 degree centrality		근접중심성 closeness centrality		사이중심성 betweenness centrality		고유벡터중심성 eigenvector centrality	
물체	141908	주기	1	속력	39436.8	속력	1
속력	87604	알짜힘	1	그림	8819.1	수레	0.639776
운동	80214	진폭	1	운동	8760.1	운동하다	0.607363
작용하다	72085	진동수	1	일정하다	6653.7	쇠구슬	0.598351
방향	65047	동영상	1	실	4898.1	진자	0.481261
그림	38971	탄성체	1	거리	4448.8	거리	0.471786
변하다	35403	말	1	수레	4082.0	이동하다	0.45215
시간	31656	썰매	1	달리다	3999.4	직선	0.445692
용수철	28640	표현하다	1	떨어지다	3970.8	구간	0.427356
일정하다	28075	영향	1	추	3660.2	주기	0.42636
나타내다	24300	무중력	1	지구	3420.7	실	0.403043
크기	24256	원운동	1	작용하다	3361.6	구슬	0.394905
합력	18821	인력	1	운동하다	3221.7	빛면	0.383048
크다	18213	자연	1	쇠구슬	3058.4	간격	0.382536
사이	17467	로켓	1	중력	2799.7	이동거리	0.381507
질량	17419	스케이트	1	뉴턴	2641.9	초	0.372217
추	17088	건물	1	합력	2233.8	타점	0.371974
공	16826	자기부상열차	1	자석	1915.7	일정하다	0.363548
중력	16751	경주	1	버스	1812.7	관성	0.357454
빠르다	16477	올림픽	1	저울	1671.0	알짜힘	0.353561

하도록, 즉, 글쓰기를 유도할 수 있다. 예를 들면 ‘힘’에 관한 글쓰기를 하되 근접중심성 상위 목록에서 ‘썰매’를 선택, ‘썰매’의 상황을 제시하고, 사이중심성 상위 목록에서 ‘속력’을 선택하여 필수 단어로 지정한다. 이 때 ‘힘’에 대한 글쓰기이므로 ‘힘’ 역시 필수 단어로 지정할 수 있다. 학생들은 ‘힘’, ‘썰매’, ‘속력’이 함께 들어갈 수 있는 상황을 떠올린 뒤 상황을 기술하는 과정에서 고유벡터중심성 상위 목록의 단어와 연결중심성 상위 목록으로부터 각각 한 개씩의 단어를 선택하여 기술할 수 있다. 필수 단어는 교사가 선정해서 제시하고 참조 단어는 학생들에게 후보 목록 중에서 각각 하나씩 선택하도록 하는 것도 방법이 될 수 있다. 아래는 학생들이 답으로 기술할 수 있는 문장의 예시를 제시한 것이다.

- ㄱ. 오른쪽으로 운동하고 있는 썰매에 오른쪽으로 힘을 작용하면 썰매의 속력이 빨라진다.
- ㄷ. 썰매에 힘이 작용하더라도 알짜힘이 0이면 속력이 일정한 운동을 하게 된다.

두 문장 모두 필수 단어는 ‘힘’, ‘썰매’, ‘속력’을 필수 단어로 포함하고 있다. ㄱ의 경우는 썰매를 밀어서 빨라지고 있는 상황을 기술하기 위하여 고유벡터중심성 상위 목록에서 ‘운동하다’를 선택하고 연결중심성 상위 목록에서 ‘작용하다’를 선택한 것이다. ㄷ의 경우는 썰매를 밀고는 있으나 움직이지 않는 상황을 기술하기 위하여 고유벡터중심성 목록에서 ‘알짜힘’을 선택하고, 연결중심성 상위 목록에서 ‘운동’을 선택한 예시이다. 과학자나 과학교육전문가 혹은 과학교사들에게는 단어를 제시하지 않더라도 ㄱ이나 ㄷ과 같이 기술하는 것이 매우 쉬운 일일 수 있다. 그러나 과학적 언어에 친숙하지 않은 학생들은 썰매를 밀어서 썰매가 빨라지고 있는 상황을 떠올린다고 하더라도 그것을 ㄱ과 같이 물리적으로 기술하는 것이 결코 쉽지 않을 것이다. 따라서 제시한 바와 같이 소재와 필수 단어, 그리고 참조 단어를 제시하여 현상을 물리적으로 기술하는 연습이 도움이 될 수 있을 것이다. 그리고 이 때 제시할 단어를 선택하는 데 있어 텍스트 네트워크 분석 결과값들을 활용하면 학생들이 쉽고 정확하게, 그리고 효율적으로 과학적 언어에 익숙해질 것으로 기대할 수 있다. 네트워크 지표에서 중심성이 높은 단어들 목록은 문장 활용성이 높은 단어들에 대한 정보이자 힘과 관련된 글에서 중심을 차지하고 있는 단어들에 대한 정보를 제공하기 때문이다.

4. 워드임베딩

가. 분석 결과

워드임베딩은 단어가 가진 의미나 속성들을 벡터값으로 환산해

주므로, 텍스트 내의 문맥을 통해 의미적, 표현적으로 유사한 단어들을 추출하는 데 유용하게 활용될 수 있다. 표 6은 워드임베딩을 수행한 뒤 ‘힘’과 유사도가 높은 상위 20개의 단어를 추출한 뒤 추출된 단어들을 ‘힘’과의 관련성을 고려하여 몇 가지로 구분해본 것이다. 단어들은 크게 힘과 개념 위계상 유사한 단어들, 그리고 힘을 구성하고 있거나 표현하기 위해 필요한 하위 요소들, 그리고 힘의 작용과 관련된 서술어들, 그리고 기타로 구분하였다.

나. 워드임베딩을 이용한 개념 학습 중요 단어 목록 작성

표 6의 결과를 앞의 결과들과 비교했을 때 특징적인 것은 다른 방법에서 중요 단어로 많이 추출되었던 ‘물체’가 워드임베딩 결과에서는 상위 20개 안에 추출되지 않았다는 점, 그리고 다른 방법들에 비해 ‘힘’의 물리적 개념을 매우 잘 드러내고 있다는 점이다. 따라서 워드임베딩 결과는 ‘힘’의 개념과 관련된 중요한 단어 목록을 작성하는 데 유용하게 활용될 수 있을 것이다. 물론 표 6의 목록이 완벽하다고 볼 수는 없다. 예를 들어 표 6에서 힘의 하위 요소에는 ‘작용점’이 빠져 있다. 워드임베딩 결과는 입력된 텍스트의 특성에 의존하므로 과학교과서 힘과 운동 단원에서 힘을 설명하는 텍스트 전반에 걸쳐 ‘작용점’이 중요하게 다루어지지 않았을 가능성이 높다. 이러한 점은 워드임베딩이라는 경험적 방법을 통해 표 6과 같은 단어 목록을 얻은 뒤, 이를 바탕으로 하여 전문가들이 연역적 방법으로 보완할 수 있겠다. 이러한 중요 단어 목록은 과학교과서 등에 학습 가이드로 제시할 수도 있고, 혹은 교육과정 성취기준에 중요 단어 목록으로 제시해줄 수도 있겠다.

IV. 결론 및 제언

본 연구에서는 과학교과서 텍스트를 계량 언어학적 방법을 적용하여 분석해 보고, 그 결과물을 하나하나 살펴 과학교육학적 의미를 탐색하여 과학교육 현장에 적용할 수 있는 방안들을 제안해 보았다. 중학교 1학년 과학교과서 ‘힘과 운동’ 단원 텍스트를 클러스터 분석, 공기 빈도 분석, 텍스트 네트워크 분석, 그리고 워드임베딩의 네 가지 계량 언어학적 분석 방법을 사용하여 분석해 보았다. 연구 결과 사용된 네 가지 텍스트 분석 기법들의 결과들은 모두 상이한 단어 목록을 보여 주었고, 네 가지 방법 각각이 서로 다른 특징과 나름의 의미를 구현하고 있음을 확인하였다. 이는 네 가지 모두 과학 교육에 유용하게 활용될 수 있으며, 교육적 의도와 목표에 따라 선택적으로 사용될 수 있음을 의미한다. 연구 결과를 요약해보면 다음과 같다. 먼저, 클러스터 분석 결과를 활용하여 문장 완성하기 활동을 제안하였다. 문장 완성하기 활동은 특정 개념에 대한 학습의 초기 단계에 초보적인 수준에서 점차로 확장시키며 과학적 문장을 작성하는 훈련으로 활용될

Table 6. Words around the ‘force’ from word embedding

구분	단어
힘의 유사어	힘이크기, 알짜힘, 합력, 중력, 마찰력, 탄성력, 자기력
힘의 하위 요소	크기, 방향, 뉴턴, 화살표
서술어	작용하다, 주다, 나란하다, 밀다, 당기다, 변하다
기타	변형, 효과, 힘의합성

수 있다. 다음으로, 공기 빈도 분석 결과를 이용한 빈 칸 채우기 활동을 제안하였다. 이는 어느 정도 기본적인 개념을 익힌 뒤 다양한 문장에 활용해 보며 용어의 용법을 익히고 개념을 정교화하는 과정에서 활용할 수 있다. 세 번째로 네트워크 분석 결과를 이용하여 소재 중심 글쓰기 활동을 제안하였다. 앞의 두 방법이 개념을 담은 용어의 형태적 사용법에 초점이 맞추어져 있다면, 네트워크 분석 결과는 실제 다양한 물리적 상황에 학습한 개념을 적용해 보고 용어의 의미적 사용법을 익히는 데 보다 도움이 될 수 있다. 이 때 학생들은 문장으로 기술할 수도 있고 문단 단위로 기술할 수도 있다. 끝으로 워드임베딩을 이용한 학습 중요 단어 목록 작성을 제안하였다. 특정 과학 개념과 관련된 중요 단어들은 교수자와 학습자 모두에게 중요한 자료가 될 수 있다. 이 단어들은 특정 개념과 의미적으로 매우 가까운 관계를 가지므로 과학의 지식 체계 속에서 해당 개념의 위상을 정교화 하고 머리 속에서 개념들 사이의 연결을 정확하게 구성하는 데 도움이 될 수 있다. 이러한 연구 결과를 바탕으로 본 연구의 결과가 과학 교육 분야에 시사할 수 있는 점을 몇 가지 아래에 제시해 보았다.

첫째, 과학교육에서 과학 언어나 과학용어에 대한 교육을 실시할 수 있는 몇 가지 구체적인 방안들을 제안하였다. 그리고 각각의 방안들을 실행하기 위한 계량 언어학적 도구들도 함께 제시하였다. 본 연구에서 제안한 방법들은 완성된 교수법이 아닌 언어학적 교수법을 도입하는 하나의 예시와 가능성을 보여주었다는 것에 중요한 의미가 있다. 지금까지 과학 교육에서 과학 언어나 과학용어에 대한 교육이 구체적으로 잘 다루어지지 않았으므로 본 연구에서 제시된 방법들이 과학 교육 분야의 다른 연구자들이나 혹은 교사들에게 다소 생소하게 느껴질 수도 있다. 그러나 과학 언어나 과학 용어의 교육의 중요성과 필요성을 고려한다면(Halliday, 1993; Rupley & Slough, 2010) 언어학적 접근 방법이나 언어학 분야의 오랜 연구 성과들을 적극적으로 활용하는 것이 필요하다고 여겨진다(Yager, 1983). 그리고 이러한 언어학적 접근 방법들이 과학 교육 현장에 적용되어 실효를 거두기 위해서는 본 연구와 같이 도입 가능성을 탐색하고 전략을 제안하는 시행착오적 연구들이 꾸준히 시도되고 검증되어야 할 것이다.

둘째, 본 연구는 경험적 언어학을 기반으로 하고 있다. 그러나 언어학 분야에서도 경험적 언어학과 연역적 언어학의 상호 보완적 역할을 강조하고 있는 것과 같이(Kwon, 2008), 과학 교육에서도 언어학의 방법을 도입하기 위해서는 본 연구와 같은 경험적 연구와 함께 연역적 연구들이 상호보완적으로 이루어져야 할 필요가 있다. 텍스트를 계량적으로 분석한 결과물들은 분명 연역적인 사고와 논의만으로는 결코 알아낼 수 없는 자료들을 제공해 주지만, 이들이 바탕으로 하고 있는 사회적 언어 결과물들, 즉, 분석 대상 텍스트의 불완전성과 변화 가능성을 필연적으로 내포하고 있기 때문에 연역적 보완이 필요하다. 본 연구에서 제시된 결과들을 예로 들어 보면, ‘힘’이 문장에서 사용될 때 함께 연이어 사용할 단어로 ‘작용하다’와 ‘물체’ 가운데 어느 것을 먼저 하는 것이 더 효과적일지, 혹은 어느 것이 학생들에게 더 쉬울지는 전문가들이 아무리 함께 논의해도 알아내기 어려울 것이다. 그러나 클러스터 분석을 통해 ‘힘 + 작용하다’의 어절 연쇄가 ‘힘 + 물체’의 어절 연쇄에 비해 6배 가까이 많이 사용된다는 것을 확인할 수 있었고, 이는 ‘힘 + 작용하다’의 어절 연쇄를 먼저 학습하는 것이 훨씬 더 쉽고 다양한 문장을 만들어낼 수 있음을 의미한다. 반면에 워드임베딩 분석 결과에서 힘의 하위 요소에 해당하는 유사어 가운데

‘크기’, ‘방향’은 추출이 되었으나 힘의 요소 가운데 하나인 ‘작용점’은 추출되지 않았다. 따라서 이러한 부분은 과학의 지식 체계 혹은 개념 체계들을 고려하여 전문가들에 의해 연역적으로 보완되어야 한다.

셋째, 본 연구에서 제안한 몇 가지 교육 방법들에 대한 효과를 검증하고 개선해 나가는 후속 연구가 이어져야 한다. 학생들의 머리 속에 과학의 지식 체계가 잘 조직되고 구성되기 위해서는 분명히 과학 언어에 충분히 노출되고 사용하는 활동이 중요하다(Shaw, 2002; Vygotsky, 1962). 그러나 본 연구에서 제안한 활동들이 실제로 어떤 학생들에게 어떤 효과가 어느 정도로 나타날 것인지 등에 대해서는 현장 적용 연구를 통해 하나 하나 검증해 보아야 한다. 아울러 언어로 표현되는 과학적 상황 가운데는 일상 생활에서 접하기 어려운 경우가 많은 데, 이런 경우 그림이나 영상과 연계한 교육 방안 등을 모색하여 교육적 효과를 높이는 전략도 탐색해볼 수 있을 것이다. 또 물리의 다른 개념들이나 물리 이외의 생명과학, 지구과학, 화학 분야 등에 적용해 보고, 학문적 특성과 개념적 특성들이 텍스트 분석 결과나 교육 활동의 효과에 어떤 특이점이 나타나는지에 대한 연구 또한 이루어져야 할 것이다.

국문요약

과학 개념에 대한 이해를 언어학적 관점에서 바라보면 학생들이 과학 개념어에 대한 깊고 정교한 이해와 더불어 정확하게 사용할 수 있는 능력을 길러주는 것이 매우 중요하다. 본 연구에서는 지금까지 과학 교육에서 과학 개념어에 대한 생산적 지식 교육의 기틀이 잘 마련되어 있지 않음에 주목하고, 과학 개념을 구성하고 있는 단어들 사이의 관계를 생산적이고 효과적으로 교육할 수 있는 방안을 탐색함으로써 과학 개념어의 생산적 지식 교육의 기틀을 제공하고자 하였다. 이를 위해 첫째, 몇 가지의 계량 언어학적 텍스트 분석 방법을 이용하여 과학 교과서 텍스트로 부터 과학 개념을 구성하고 있는 단어들과 그들 사이의 관계를 추출하고, 둘째, 각 방법의 결과로 추출된 단어 관계의 의미를 정성적으로 살펴본 뒤, 셋째, 이를 이용하여 과학 개념어의 생산적 지식 향상에 도움을 줄 수 있는 쓰기 활동 방법을 제안해 보았다. 중학교 1학년 과학교과서 ‘힘과 운동’ 단원 텍스트를 클러스터 분석, 공기 빈도 분석, 텍스트 네트워크 분석, 그리고 워드임베딩의 네 가지 계량 언어학적 분석 방법을 사용하여 분석해 보았다. 연구 결과 첫째, 클러스터 분석 결과를 활용하여 문장 완성하기 활동을 제안하였다. 둘째, 공기 빈도 분석 결과를 이용한 빈 칸 채우기 활동을 제안하였다. 셋째, 네트워크 분석 결과를 이용하여 소재 중심 글쓰기 활동을 제안하였다. 넷째, 워드임베딩을 이용한 학습 중요 단어 목록 작성을 제안하였다.

: 과학 개념어, 과학 개념, 과학 글쓰기, 클러스터분석, 공기 빈도 분석, 텍스트 네트워크 분석, 워드임베딩

References

- Barcroft, J. (2004). Effects of sentence writing in second language lexical acquisition. *Second Language Research*, 20, 303-334.
- Barnbrook, G. (1996). *Language and computers*, Edinburgh University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad, & E. Finegan. (1999). *Longman Grammar of spoken and written English*. London: Longman.
- Choi, S., Mattesson, A. S., & Lim, H. (2018). Utilizing local bilingual

- embeddings on Korean-English law data. *Journal of the Korea Convergence Society*, 9(10), 45-53.
- Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Lawrence Erlbaum Associates Publishers.
- Folse, K. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, 40(2), 273-293.
- Griffin, G. F., & Harely, T. A. (1996). List learning of second language vocabulary. *Applied Psycholinguistics*, 17(4), 443-460.
- Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5, 93-116.
- Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, 51, 539-558.
- Jhang, S., & Lee, S. (2013). Clusters and key clusters in the Maritime English Corpus. *Journal of Language Sciences*, 20(4), 199-219.
- Kang, B. (2011). *Language, computer, corpus linguistics*. Korea University Press.
- Kim, J. & Im, H. (2012). The interaction effects of exposure frequency and task-induced involvement on English learners' vocabulary development. *Foreign Languages Education*, 19(4), 349-373.
- Kwon, H. (2008). The practice and application of corpus linguistics. *Korean Journal of Applied Linguistics*, 24(3), 1-30.
- Laufer, B. (1998). The development of passive and active vocabulary: Same or different? *Applied Linguistics*, 19(2), 255-271.
- Lee, I., & Kim, M. (2006). Multi-document summarization based on cluster using term co-occurrence. *Journal of KIISE: Software and Applications*, 33(2), 243-251.
- Lee, C. (1998). Basic study for development of Korean vocabulary teaching theory. *Korean Language Education*, 98, 75-103.
- Mikolove, T., Sutskever, I., Chen, K., Corrado, GS., & Dean, J. (2013a). Distributed representations of words and phrases and their compositionality. *Neural Information Processing System*, 26, 3111-3119.
- Nation, I. S. P. (1990). *Learning and teaching vocabulary*. New York: Newbury House.
- Nattinger, J. (1988). Some current trends in vocabulary teaching. In R. Carter & M. McCarthy (Eds), *Vocabulary and language teaching* (pp. 62-82). London: Longman.
- Park, B. (2002). A study on Korean syntactic collocation. *The Studies of Korean Language and Literature*, 19, 333-351.
- Piaget, J. (1926). *The language and thought of the child*. Harcourt, Brace.
- Rincke, K. (2011). It's rather like learning a language: Development of talk and conceptual understanding in mechanics lessons. *International Journal of Science Education*, 33(2), 229-258.
- Robinson, P. (2008). *Handbook of cognitive linguistics and second language acquisition*. Routledge.
- Rupley, W., & Slough, S. (2010). Building prior knowledge and vocabulary in science in the intermediate grades: Creating hooks for learning. *Literacy Research and Instruction*, 49, 99-112.
- Ryoo, Y. (2009). Effects of two types of vocabulary practice: Receptive and productive. *Foreign Languages Education*, 16(1), 79-99.
- Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Disability Studies*, 20, 33-53.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Shaw, J. (2002). Linguistically responsive science teaching. *Electronic Magazine of Multicultural Education*, 4(1), 24.
- Shin, J. C., & Ock, C. Y. (2012). A stage transition model for Korean part-of-speech and homograph tagging. *Software and Application*, 39(11), 889-901.
- Shin, W., Kim, I., & Kim, H. (2010). A study of spatial nouns and network analysis based on corpus. *Text Linguistics*, 29, 221-250.
- Shintani, N. (2011). A comparative study of the effects of input-based and production-based instruction on vocabulary acquisition by young EFL learners. *Language Teaching Research*, 15, 137-158.
- Stubbs, M. (1995). Collocations and semantic profiles. *Functions of Language*, 2, 1.
- Vygotsky, L. S. (1962). *Thought and language*. MA: MIT Press.
- Waring, R. (1997). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature*, 21(1), 94-114.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Wittgenstein, L. (1958). *The blue and brown books*. Oxford: Blackwell.
- Yager, R. (1983). The importance of terminology in teaching K-12 science. *Journal of Research in Science Teaching*, 20, 577-588.
- Yun, E. & Park, Y. (2016). Comparison of the usage of the science terminology 'force' in science and everyday languages. *New Physics: Sae Mulli*, 66(4), 460-467.
- Yun, E. & Park, Y. (2018). Extraction of scientific semantic networks from science textbooks and comparison with science teachers' spoken language by text network analysis. *International Journal of Science Education*, 40(17), 2118-2136.
- Yun, E., Kwon, S. G., & Park, Y. (2015). Analysis of problems of current science textbooks perceived by teachers and students in view of learner-centered classroom. *Journal of Science Education*, 39(3), 404-417.
- Zhou, S. (2010). Comparing receptive and productive academic vocabulary knowledge of Chinese EFL learners. *Asian Social Science*, 6(10), 14-19.

저자 정보

윤은정(경북대학교과학교육연구소 초빙교수)