



과학교과서의 학년 간 언어적 특성 분석 -텍스트 정합성을 중심으로-

류지수, 전문기*
건국대학교

An Analysis of Linguistic Features in Science Textbooks across Grade Levels: Focus on Text Cohesion

Jisu Ryu, Moongee Jeon*
Konkuk University

ARTICLE INFO

Article history:

Received 26 January 2021

Received in revised form

16 February 2021

8 April 2021

16 April 2021

Accepted 19 April 2021

Keywords:

science textbooks, text analysis,
Korean text analysis,
text cohesion

ABSTRACT

Learning efficiency can be maximized by careful matching of text features to expected reader features (i.e., linguistic and cognitive abilities, and background knowledge). The present study aims to explore whether this systematic principle is reflected in the development of science textbooks. The current study examined science textbook texts on 20 measures provided by Auto-Kohesion, a Korean language analysis tool. In addition to surface-level features (basic counts, word-related measures, syntactic complexity measures) which have been commonly used in previous text analysis studies, the present study included cohesion-related features as well (noun overlap ratios, connectives, pronouns). The main findings demonstrate that the surface measures (e.g., word and sentence length, word frequency) overall increased in complexity with grade levels, whereas the majority of the other measures, particularly cohesion-related measures, did not systematically vary across grade levels. The current results suggest that students of lower grades are expected to experience learning difficulties and lowered motivation due to the challenging texts. Textbooks are also not likely to be suitable for students of higher grades to develop the ability to process difficulty level texts required for higher education. The current study suggests that various text-related features including cohesion-related measures need to be carefully considered in the process of textbook development.

1. 서론

1. 과학교과서의 중요성

교과서는 교육과정의 목표를 달성하기 위하여 개발된 학습 자료로써, 교수 및 학습에 있어서 가장 핵심적인 학습 매개체이다(Ryu & Jeon, 2020). 특히 한국과 같이 국정 및 검인정 제도에 따라서 교과서가 개발되는 경우, 교과서는 학습 시 절대적인 정보 매개체로 인식되어 그 중요성이 보다 강조된다(Kim, Park, Cho, & Song, 2014). 그중에서도 국어와 같은 다른 교과서에 비하면 과학교과서에 대한 학습자의 의존도는 매우 높다(Koh, Song, & Kang, 2010). 이러한 이유는 과학적 원리와 지식체계에 대한 대부분의 학습 기회와 노출이 다른 분야에 비하면 상대적으로 학교에서 사용되는 교과서에 국한되어 있는 것과 연관된다. 이처럼 교과서가 교수 및 학습에 미치는 영향이 매우 크기 때문에 교과서의 적절성에 대한 문제는 중대한 교육적 함의를 내포한다(Chall & Conrad, 1991; Gasparinatou & Grigoriadou, 2013; Voss & Silfies, 1996). 따라서 교과서의 적절성에 대한 평가는 매우 중요한 문제이다. 이러한 교과서의 적절성을 평가하는 데 중요하게 고려해야 할 사항은 교과서에 반영된 다양한 언어적 특성

(linguistic characteristics)이다(Ryu & Jeon, 2020). 왜냐하면, 교과서의 많은 내용이 문자 언어, 즉 텍스트 형태로 학습자에게 전달된다는 사실을 고려할 때, 학습자의 정보 습득 및 학습은 텍스트의 특성에 상당 부분 결정될 수 있기 때문이다(Lee, 2011). 이러한 맥락에서, 독자의 인지 발달 수준에 따라 충분히 이해될 수 있는 형태로 교과서 텍스트가 집필되어 있는지, 그리하여 교과서를 통한 학습 효율이 최대화될 수 있는 형태로 교과서 텍스트가 설계되어 있는지에 대해 연구자들의 관심이 집중되었다(e.g., Berendes *et al.*, 2018).

요약하면, 텍스트를 통한 정보 습득 및 학습은 기본적으로 텍스트를 성공적으로 이해하는 것에서 기인한다(Vellutino, 2003). 따라서 성공적인 정보 습득과 학습을 위해 교과서의 다양한 특성을 분석하는 것이 중요하다(Ryu & Jeon, 2020). 본 연구에서는 중학교 과학 텍스트의 학년 간 언어적 특성을 비교 분석하고자 하였다. 이러한 본 연구의 목적을 위해 먼저 과학 텍스트를 성공적으로 이해하는 데 영향을 미치는 여러 요인에 대해 개괄하고자 한다. 또한, 과학교과서에 수록된 텍스트의 다양한 언어적 특성을 분석한 연구를 소개하고자 한다.

2. 과학교과서의 텍스트 이해에 미치는 요인

텍스트를 통해 학습이 일어나는 과정을 설명하는 이론적 접근 중

* 교신저자 : 전문기 (mjeon1@konkuk.ac.kr)

<http://dx.doi.org/10.14697/jkase.2021.41.2.71>

가장 널리 받아들여지는 것은 구성주의(constructivism) 접근이다 (Graesser & Clark, 1985; van Dijk & Kintsch, 1983). 이 이론에서는 텍스트와 독자 간의 상호작용(interaction)을 강조한다. 다시 말하면, 구성주의 접근은 독자가 텍스트를 통해 명시적으로 얻을 수 있는 정보를 자신의 기존 인지체계와 통합시키는 데 성공해야만 그 텍스트를 이해할 수 있고, 그로 인해 학습이 성공적으로 일어난다는 것을 전제한다(Beck, McKeown, Sinatra, & Loxterman, 1991). 예를 들어, Kintsch(1988, 1998)의 구성-통합 모형(Construction-Integration Model)에 따르면, 텍스트를 이해하기 위해 독자는 먼저 자신의 언어 지식을 활용하여 텍스트에 명시적으로 제시되어 있는 정보를 바탕으로 심적 표상(mental representation)을 형성한다. 이 단계에서는 텍스트가 전달하는 정보에 대한 표상이 텍스트에 명시적으로 주어진 순수한 언어적 특성에 의해서만 형성되기 때문에 텍스트에 대한 표상이 완전하지 못하다. 왜냐하면, 텍스트에 대한 이해는 텍스트의 명시적인 정보뿐만 아니라 이러한 명시적인 정보와 이해자의 사전지식(prior knowledge) 혹은 배경지식(background knowledge)과의 능동적인 상호작용에 기반해 일어나기 때문이다(Kintsch, 1998). 다시 말하면, 텍스트에 명시적으로 제시되는 정보들만으로 텍스트를 성공적으로 이해하는 데 어려움이 수반되기 때문에 텍스트를 통해 얻은 정보들을 하나의 일관된 표상으로 유기적으로 연결시키기 위해서는 이해자의 사전지식 혹은 배경지식을 바탕으로 명시적으로 언급되지 않은 부분들을 추론(inference)해야 하는 과정이 요구된다. 이러한 추론과정에 독자의 사전지식 혹은 배경지식이 활용되는 것이다. 결과적으로 텍스트에 기반한 정보(text-based information)에 배경지식이 적절히 통합되면 텍스트 내 문장 간의 응집성(coherence)이 높아지고 전체적으로 유기적이고 일관된 심적 표상을 형성할 수 있게 된다(Kintsch, 1998). 이러한 심적 표상 과정은 학습 과정과 긴밀하게 연결되어 있다(McNamara, Graesser, & Louwerse, 2009). 구체적으로 제시하면, 독자가 텍스트에 대한 일관된 표상을 형성하는데 추론이 많이 요구될수록 그에 상응하게 배경지식 또한 많이 요구되기 때문에 학습 효과가 극대화될 수 있다(Best, Rowe, Ozuru, & McNamara, 2005). 반면, 텍스트 내 문장들 간의 관계가 명시적인 텍스트 기반 정보에 크게 의존해 그 텍스트 정보만으로도 충분히 짜임새 있게 구성된다면 독자가 문장 간 관계를 파악하기 위해 많은 추론을 할 필요가 없기 때문에 결과적으로 충분한 학습 효과가 나타나지 않을 수 있다. 왜냐하면, 학습자가 이해하고자 하는 텍스트에 대해 추론을 많이 할수록 학습을 통해 얻는 것이 많기 때문이다(McNamara et al., 2009). 결국, 구성주의 접근은 텍스트에 의한 독자의 학습 효과를 극대화하기 위해 독자 요인과 텍스트 요인 모두를 고려하는 것이 중요함을 제시한다.

이러한 맥락에서 먼저 텍스트 이해에 영향을 미치는 독자 요인에 대해 많은 연구가 수행되었는데 그중 가장 영향력이 크고 지속적으로 연구가 수행된 요인은 독자가 특정 분야에 대해 갖고 있는 배경 지식(background knowledge)이다(Best et al., 2005; Boscolo & Mason, 2003; Gilabert, Martinez, & Vidal-Abarca, 2005; Kendeou & van den Broek, 2007; McNamara, 2001; O'Reilly & McNamara, 2007; Voss & Silfies, 1996). 많은 연구들은 독자가 새로운 정보를 이해하고 학습하는 데 있어 배경지식이 매우 큰 영향을 미친다는 결과들을 제시하였다. 전술한 것처럼, 텍스트를 통해 정보가 성공적으로 습득되기 위해서는 텍스트에서 추출한 명시적인 정보가 독자의 기존 인지

구조에 적절하게 통합되어야 하기 때문에 독자의 기존 배경 지식은 새로운 정보를 학습하기 위한 토대로써 기능한다(O'Reilly & McNamara, 2007). 그렇기 때문에, 해당 글에 대한 배경 지식이 많은 독자는 텍스트에 기반한 정보를 기존 인지 체계에 보다 잘 통합시킬 수 있다. 따라서 이러한 독자는 텍스트 내 문장 간의 관계를 보다 분명하게 파악할 수 있고 내용도 잘 이해할 수 있다(Best et al., 2005; Chiesi, Spilich, & Voss, 1979; McKeown, Beck, Sinatra, & Loxterman, 1992; Pressley et al., 1992; Voss & Silfies, 1996).

다음으로, 텍스트 이해에 영향을 미치는 텍스트 관련 요인 중 핵심적인 요인은 텍스트 구조이다. 독자의 이해 과정은 텍스트에 제시된 언어적 구성 요소(linguistic factors)의 사용 정도에 의해 크게 영향을 받는다(Kendeou & van den Broek, 2007). 따라서 내용이 유사한 두 개의 텍스트에 대한 이해도는 각각의 텍스트에 포함된 언어적 구성 요소에 의해 크게 달라질 수 있다. 이러한 언어적 구성 요소는 텍스트의 정합성(cohesion) 형성에 직접적인 영향을 미치는 요인으로써 활발히 연구되고 있다. 정합성이란 텍스트 내에 명시적으로 제시된 언어적 요소들에 기반해서 형성된 텍스트에 대한 표상을 지칭한다(O'Reilly & McNamara, 2007). 이러한 정합성은 텍스트 내에 있는 여러 언어적 장치들(e.g., 접속어, 논항 반복, 대명사, 어휘 반복, 동의어, 상위 개념어, 대치, 생략 등)을 통해 조절 및 형성될 수 있다(Halliday & Hasan, 1976). 반면에, 정합성보다 좀 더 확장된 의미를 내포하고 있는 응집성(coherence)은 텍스트에 기반한 정보(text-based information)와 이해자의 배경지식(background knowledge)과의 상호작용에 의해서 형성된 텍스트에 대한 표상이다(Graesser, Jeon, Yan, & Cai, 2007). 이 두 용어는 학문 분야에 따라서 달리 사용될 수 있으나 본 연구에서는 텍스트에 기반해 형성된 텍스트에 대한 표상을 정합성으로 지칭한다. 따라서 텍스트 내에 명시적으로 제시된 여러 언어적 장치들(linguistic devices)에 의해서 텍스트 내의 문장들이 서로 긴밀하게 연결이 되어 제시된다면, 이 텍스트의 정합성은 높다고 할 수 있다(Best et al., 2005). 반면, 상대적으로 이러한 언어적 장치들이 부재하여 문장 간의 관계가 명시적으로 파악되지 않는 텍스트의 정합성은 낮다고 볼 수 있다. 일반적으로 독자는 정합성이 낮은 텍스트를 읽을 때, 명시적으로 제시되지 않은 관계를 파악하기 위해 많은 추론을 하는 경향이 있다(Best et al., 2005). 구체적으로, Best et al.(2005)의 연구에 제시된 아래 예시 문장들을 통해 언어적 요소들이 정합성 형성에 미치는 영향을 설명하면 다음과 같다.

(1) 우리는 유성 생식의 존재를 당연하게 여기는 경향이 있다. 진화론적 관점에서 보면, *이것*은 심각한 난제이다.

(2) 우리는 유성 생식의 존재를 당연하게 여기는 경향이 있다. *하지만*, 진화론적 관점에서 보면, *유성 생식*은 심각한 난제이다.

두 예시문은 상당히 유사한 내용을 다루고 있지만 두 번째 예시문이 첫 번째 예시문보다 더 정합성이 높다. 먼저, 첫 번째 예시문에는 '이것'이라는 지시 대명사가 사용되었는데, 독자는 '이것'이 가리키는 지시대상(referent)을 찾기 위해 추론을 해야 한다. 언어적 관점에서, '이것'은 '유성 생식'을 가리킬 수도 있고 '우리는 유성 생식의 존재를 당연하게 여기는 경향이 있다.'라는 문장 전체를 가리킬 수도 있다. 따라서, 두 문장의 관계를 파악하기 위해 독자의 배경 지식이 상대적으로 많이 요구된다. 반면, 두 번째 예시문에는 대명사 대신에 '유성 생식'이라는 명사구가 반복되어 사용되고 있다. 따라서, 두 문장을

연결시키기 위한 추론 과정의 필요성이 경감된다. 또한, 두 번째 예시문에 사용된 언어적 단서(linguistic cue)인 ‘하지만’이라는 접속부사는 두 문장이 역접의 관계로 연결된다는 것을 명시적으로 보여주고 있다. 이러한 예시문들을 통해 명사구 반복, 접속 부사와 같은 명시적인 언어적 연결 장치들이 텍스트의 정합성 형성에 중요한 영향을 미침을 알 수 있다.

요약하면, 텍스트를 이해하여 학습 효과가 나타나는 것은 독자가 사전에 갖고 있던 배경지식과 같은 독자 관련 요인과 텍스트에 기반한 명시적인 언어적 정보와 같은 텍스트 관련 요인이 상호작용한 결과에서 기인한다고 할 수 있다. 또한, 독자 요인과 텍스트 요인은 본질적으로 텍스트의 응집성(coherence) 형성에 영향을 미치는 서로 다른 언어적 구성(linguistic construct)개념이지만 상호 보완적인 성격을 지니고 있다. 정합성이 낮은 글은 독자가 많은 추론을 통해 문장 간 관계를 파악해야 하므로 독자의 배경지식 의존도를 높이지만, 정합성이 높은 글은 문장 간의 관계가 많은 부분 이미 텍스트 내에 명시적으로 제시된 언어적 장치들을 통해 파악되기 때문에 독자의 배경지식 의존도를 낮추는데 기여한다(Ozuru, Dempsey, & McNamara, 2009).

이러한 맥락에서, 특히 과학교과서와 같은 복잡한 내용을 담은 설명문(expository text)과 유사한 성격의 텍스트에서는 정합성의 역할이 보다 중요해진다. 과학 교과목의 특성을 고려해볼 때 대부분 충분한 배경지식이 없는 독자들이 많기 때문에 텍스트 구조상의 비계(scaffolding) 설정이 필요하다(Vygotsky, 1978). 다시 말하면, 독자의 이해 수준에 부합되는 텍스트가 제공될 필요가 있다. 학습 효과에 대한 정합성의 긍정적 영향에 주목한 연구자들은 과학 텍스트의 정합성을 증가시킨 결과(e.g., 대명사 대신 명사구 중복, 연결어 사용 등의 방식을 활용), 과학 텍스트에 대한 배경 지식이 부족한 학습자들의 이해도가 상당히 증가한 것을 확인하였다(Gasparinatou & Grigoriadou, 2011; Gilibert *et al.*, 2005; Hall, Maltby, Filik, & Paterson, 2016; McNamara, 2001; Ozuru *et al.*, 2009; Ozuru, Dempsey, Sayroo, & McNamara, 2005; Silfhout, Everts-Vermeul & Sanders, 2014).

이처럼 정합성이 과학 텍스트 이해에 작용하는 긍정적 효과를 고려하면, 정합성이 높은 텍스트가 모든 독자에게 항상 긍정적인 결과를 가져올 것이라는 직관적인 가정을 할 수 있다. 그러나, 텍스트를 이해하는 것이 텍스트 특성(i.e., 정합성 정도)과 독자의 특성(i.e., 배경지식, 사전 지식 수준)이 함께 상호작용하는 과정이라는 구성주의적 관점을 견지한다면, 정합성의 영향은 독자의 특성에 따라 달라질 것이라 기대할 수 있다(Ozuru *et al.*, 2005). 독자의 배경 지식과 텍스트 정합성의 상호작용 효과를 분석한 연구들은 이러한 가정에 대한 실증적 근거를 제시해준다. 예를 들면, Gasparinatou와 Grigoriadou (2013)는 독자의 배경지식과 정합성이 과학 텍스트 이해도에 미치는 영향을 분석하기 위해 176명의 대학생들에게 컴퓨터 공학 관련 텍스트를 읽게 하였다. 절반의 학생들은 원본의 텍스트에서 정합성이 최대치로 조절된 버전을 읽었고, 나머지 학생들은 최저치로 조절된 버전을 읽었다. 이때, 정합성을 향상시키기 위해 명사구 반복을 증가시키고, ‘however,’ ‘therefore,’ ‘because,’ 와 같은 연결어를 덧붙였다. 연구 결과, 배경 지식이 적은 학생들의 경우 정합성이 높은 버전을 읽었을 때 상당 수준까지 이해도가 향상되었다. 반면, 배경 지식이 많은 학생들의 경우에는 정합성이 높은 텍스트보다 오히려 낮은 텍스트를 읽었을 때 높은 수준의 이해력을 묻는 질문들에 대해 보다 잘 대답할 수

있었다. 이러한 결과는 배경 지식과 정합성의 상호작용을 분석한 많은 연구에서 일관되게 보고되었다(Best *et al.*, 2005; Gilibert *et al.*, 2005; McNamara, 2001; McNamara & Kintsch, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996; Ozuru *et al.*, 2005). 이러한 현상은 “정합성 역전 효과(reverse cohesion effect)”에 의해 설명될 수 있다(McNamara & Kintsch, 1996). 다시 말하면 배경 지식이 많은 독자에게 정합성이 높은 텍스트가 제시될 경우, 텍스트 내 제시된 명시적인 정보와 독자의 배경지식 간에 잉여성이 존재하게 된다(Best *et al.*, 2005). 이 때문에 독자는 텍스트에 대한 의미 표상 형성 과정에서 각각의 문장을 연결하기 위해 추론을 할 필요성이 감소하고, 따라서 능동적인 추론 과정에 기인한 학습 효과의 극대화에 낙공하여 결국 깊은 수준의 이해까지 도달하기 어려워진다. 반면, 정합성이 낮은 텍스트의 경우에는 독자가 문장 간의 관계를 연결시키기 위해 많은 추론을 하게 되는데 그 과정에서 자신의 배경 지식을 텍스트 정보와 통합시키려는 시도를 하게 된다. 이렇게 텍스트에 대한 의미 형성 과정에 추론 전략을 활용하여 능동적으로 참여하게 되면 텍스트에 제시된 정보가 훨씬 더 기억에 잘 남고, 결국 학습 효과가 향상된다(McNamara *et al.*, 2009; Voss & Silfies, 1996).

결과들은 텍스트의 정합성을 조절할 때, 학습 효율을 증대화하기 위해서는 서로 배경 지식이 다른 독자의 특성을 고려해야 함을 보여준다. 특히 전문한 것처럼, 설명문과 같이 정보 전달이 목적인 과학교과서의 경우 이러한 연구 결과들이 세심하게 반영되어 집필될 필요가 있다. 따라서, 과학 교육에 중추적 역할을 담당하고 있는 과학교과서에 이러한 연구 결과가 반영되어 있는지 분석해보는 것이 필요하다.

3. 과학교과서의 텍스트 특성 분석 연구

연구자들은 교과서에 반영된 정합성과 같은 텍스트 특성을 조절할 때, 학습자들의 인지 수준 및 배경 지식 정도를 고려해야 한다고 주장한다(Lee, 2011; Lee, Maeng, Kim, Kim, & Kim, 2007; Koh *et al.*, 2010). 또한, Kintsch(1994)의 학습 가능 영역(zones of learnability)에 따르면 텍스트에 제시된 새로운 정보가 독자의 기존 지식체계에 점진적으로 통합이 되면서 학습이 일어나기 때문에, 독자의 사전 지식과 학습 자료 간에 중첩되는 부분이 거의 없거나 거의 대부분이 중첩된다면 학습이 일어나기 어렵다. 부연하면, 학습 자료는 학습자의 현재 수준보다 조금 더 높은 수준에서 제시되어야 한다는 것을 알 수 있다. 하지만, 연구자가 독자의 특성과 같은 개인차 변인을 직접적으로 조절하는 것은 어려우므로, 교과서 연구자들은 학습자의 수준을 일정하게 가정하고 조절이 가능한 텍스트 특성을 개인 수준에 맞추어 조절하는 것에 집중하였다(Koh *et al.*, 2010; Lee *et al.*, 2007). 이때, 학습자의 수준을 가늠하는 척도가 되는 것이 학년 수준(grade level)이다. 독자의 특성을 고려하여 텍스트 특성을 효과적으로 조절하기 위해서는 각 학년에 따른 학습자의 인지 수준 및 배경 지식 등 여러 학습자 요인에 대한 객관적이고 절대적인 준거가 필요하나 현재 체계적인 기준이 부재한 실정이다(Lee, 2011). 따라서 연구자들은 대안으로써 상대적인 비교를 통해 학년이 증가함에 따라 이전 학년에 비하여 교과서의 텍스트 복잡도가 단계적으로 증가하는지를 분석하였다(Berendes *et al.*, 2018; Dufty *et al.*, 2006; Kang & Koh, 2014; McNamara *et al.*, 2009; To, 2018; Yun, 2019). 다시 말하면, 학년이

증가함에 따라 학생들의 인지 능력, 읽기 능력, 배경 지식이 증가할 것이라 기대되기 때문에(Berendes *et al.*, 2018; Koh *et al.*, 2010; Ozuru *et al.*, 2005) 이에 따라 체계적으로 텍스트 복잡도가 상승되어야 한다고 기대할 수 있다. 이와 같은 논지는 학습의 효율성을 설명하는 여러 이론적 가설에 기반하고 있다(Berendes *et al.*, 2018; Kintsch, 1994; Krashen, 1985; Pearson, 2013; Vygotsky, 1978).

예를 들면, Dufty *et al.*(2006)은 학년이 증가함에 따라 여러 정합성 측정치가 어떻게 변화하는지를 알아보기 위해 유치원부터 고등학교 전 학년에 대한 사회 및 과학교과서를 추출하여 학년 수준과 여러 정합성 측정치 간의 상관관계를 분석하였다. 그 결과, 학년이 증가함에 따라 논항 중복(argument overlap)을 포함한 여러 명시적인 언어적 장치들의 사용빈도가 낮아지는 것으로 나타났다. 다시 말하면, 학년 증가할수록 정합성이 낮은 텍스트가 사용되었다. 이러한 결과는 분석에 사용된 교과서의 텍스트 정합성 정도가 독자 변인(학습자의 인지 수준 및 배경 지식)에 상응하게 잘 조절된 것으로 이해된다. 즉, 저학년 학습자들의 인지 및 읽기 수준은 상대적으로 낮고 학습 내용에 대한 배경지식도 적을 것이다. 따라서 배경지식이 적은 학습자에게 정합성이 높은 텍스트가 제시되면, 그 학습자의 이해 및 학습이 촉진되리라 기대된다. 환언하면, 교과서가 이전 학년에서 학습한 내용을 토대로 점점 더 복잡한 내용을 다루는 방식으로 설계되어 있다는 것을 고려하면(Ozuru *et al.*, 2005), 학년이 증가함에 따라 그에 맞게 인지 및 읽기 발달이 이루어지고 배경 지식도 점진적으로 증가되기 때문에 학습자들이 상대적으로 명시적인 언어적 정보에 의존할 필요성이 줄어든다. 이러한 점진적인 독자 특성의 변화에 대응하여, 정합성이 낮은 텍스트가 학습자에게 제시하면 추론 과정이 증가하여 텍스트의 복잡도가 상승하게 된다. 따라서 정합성이 낮은 텍스트의 경우, 텍스트 내에 제시된 개별적인 정보 간의 유기적 연결을 위해 학습자의 기존 배경 지식이 좀더 적극적으로 활용되어 텍스트 정보와 통합되어야 함이 요구된다. 이러한 과정 속에서 학습자는 의미 형성 과정에 보다 능동적으로 참여하게 된다. 따라서 더 깊은 수준의 이해가 가능하고 학습의 효과가 증대될 것이라 기대된다.

그러나 국내에서 수행된 과학교과서의 학년 간 텍스트 특성을 분석한 연구는 아직까지 그 수가 많지 않다. 한 학년 내에서 교과서의 텍스트 특성을 분석한 연구는 어느 정도 수행되었지만(Kang & Park, 2009; Lee, 2011), 학년 간 텍스트 특성의 변화를 살펴본 연구는 드물다(Kang & Koh, 2014; Yun, 2019). 더욱이 텍스트 특성 분석을 위해 선행 연구에서 사용된 언어적 측정치가 주로 어휘 다양성(lexical diversity), 어휘 길이, 문장 길이와 같은 표층적인 수준(surface level)에 국한되어 있었기 때문에 정합성 형성에 크게 기여하는 논항반복, 대명사 사용빈도, 접속어, 어휘 빈도 등과 같은 보다 심층 수준의(deep-level) 언어적 특질들이 학년이 증가함에 따라 어떻게 조절되는지에 대한 이슈가 다루어지기에는 한계가 있었다. 예를 들면, Yun(2019)은 학년 간 텍스트 특성을 비교 분석하기 위해 초등학교 3학년, 중학교 1학년, 고등학교 1학년 물리 과학교과서에 포함된 물리 용어에 대한 정의문을 추출하였다. 그런 후에 추출된 정의문에 대한 문장 길이, 통사적 복잡도, 정의문에 포함된 과학 용어의 개수 등을 분석하였다. 그 결과, 정의문의 문장은 교과서의 다른 문장과 비교했을 때 평균적으로 2, 3어절 정도 더 긴 11~14 어절로 구성되어 있었다(문장 길이 분석결과). 또, 정의문의 통사구조는 세 학년에 걸쳐

유사한 수준으로 나타났다. 구체적으로 정의문들은 3~5개의 절로 이루어진 복문과 중복문 형태의 복잡한 문장으로 구성되어 있었다. 이 연구는 초등학교, 중학교, 고등학교를 포함한 세 학년의 텍스트 복잡도를 비교 분석하였다는 점에서 의의가 있지만, 특정 단원에 국한된 정의문에 대해서만 분석하고 한정된 측정치에 대해서만 분석한 점에서 제한점을 시사하였다.

선행 연구에서 제시된 이러한 한계점에 주목하면서, 현재 연구에서는 한국어 분석 프로그램인 Auto-Kohesion 시스템(Ryu & Jeon, 2020)을 활용하여 중학교 1, 2, 3학년 과학교과서의 언어적 특성을 비교 분석하고자 한다. 분석을 위해 Auto-Kohesion 시스템이 제공하는 총 20개의 다양한 언어적 및 심리언어적 측정치를 활용하고자 한다. 이를 통해 학년이 증가함에 따라 텍스트 특성이 어떻게 변화하는지 다각도로 비교 분석하여 교과서 개발 및 평가 과정에 여러 시사점을 주고자 한다.

II. 연구방법

1. 과학교과서 텍스트 분석 시스템

본 연구는 2015년에 개정된 중학교 1~3학년 과학교과서에 수록된 텍스트 자료를 전산화하여 학년 간 변화하는 언어적 특성을 비교하고자 한다. 이를 위해 한국어 텍스트를 다양한 언어적 및 심리언어적 측정치를 통해 자동으로 분석해주는 Auto-Kohesion 시스템을 활용하였다. Auto-Kohesion 시스템은 건국대학교 다언어다문화연구소에서 코퍼스언어학 및 전산언어학 분야에서 널리 이용되고 있는 컴퓨터 알고리즘 방식 및 코퍼스 자료 등을 통해 개발되었다(Ryu & Jeon, 2020). 이 시스템은 기존 연구에서 교과서 분석 및 장르 분석 등 다양한 한국어 텍스트의 언어적 특성을 분석하는 데 활용되었다(Ryu & Jeon, 2020; Ryu, Lee, & Jeon, 2020). 이러한 이유 때문에 과학교과서에 수록된 한국어 텍스트의 언어적 특성을 비교 분석하고자 한 현재 연구의 목적에도 잘 부합하리라 판단하여 Auto-Kohesion 시스템을 본 연구에 이용하였다.

Auto-Kohesion 시스템은 여러 어휘 관련 측정치 및 정합성 관련 측정치를 포함하여 총 21개의 측정치를 제공한다. 이 중 텍스트 구조 분석 측정치는 에세이 구조 분석에 활용되므로 현재 연구의 목적에는 적합하지 않다고 판단되어 제외하였다. 따라서, 현재 연구에서는 이를 제외한 총 20개의 측정치를 활용하였다. 먼저, 본 연구를 위해 기본 측정치에 해당하는 텍스트 표층구조 측정치(어절 및 문장 수, 평균 어절 길이 및 문장 길이)를 분석하였다. 또한, 어휘 정보 측정치(전체 어휘 및 내용에 대한 빈도수, 전체 어휘 및 내용에 대한 대수변환된 빈도수, 타입-토큰 비율) 및 문장의 통사구조를 분석하는 통사구조 측정치(수식어 비율, 문장 성분 비율)를 분석하였다. 마지막으로, 정합성과 직접적인 관련이 있는 측정치인 명사 반복 측정치(인접된 문장 및 전체 문장 간 명사 반복 비율), 대명사 측정치(전체, 1인칭, 2인칭, 3인칭 대명사), 연결어 측정치(전체, 인과, 부가 접속사)를 분석하였다.

한국어 과학교과서의 언어적 특성을 비교 분석한 선행 연구에서는 주로 어절 및 문장 수, 평균 어절 길이 및 문장 길이와 같은 표층 수준에서의 측정치, 어휘 정보 측정치 및 통사적 측정치를 활용하였

다(Chang & Lim, 2016; Kang & Back, 2020; Kang & Koh, 2014; Lee, 2011; Yun, 2019). 따라서, 정합성 관련 측정치를 분석한 연구는 거의 찾아보기 어려웠다. 본 연구에서는 선행 연구에서 교과서 텍스트의 언어적 특성을 분석하기 위해 널리 활용한 표층 수준에서의 측정치, 어휘 정보 측정치 및 통사적 측정치는 물론 정합성과 직접적으로 관련이 있는 측정치까지 포함하여 종합적으로 분석하였다.

2. 분석 대상 교과서 텍스트

본 연구에서는 2015년에 개정되어 현재 사용 중인 중학교 1~3학년 과학교과서에 수록된 본문 텍스트의 언어적 특성을 비교 분석하고자 한다. 본 연구에서 활용된 Auto-Kohesion 시스템은 확장자가 txt인 파일(ASCII 코드 형식)만 분석할 수 있다. 따라서 과학교과서 본문 텍스트 코퍼스(corpus)를 모두 txt 파일로 구축하였다. 하나의 주제로 구성된 단락을 기준으로 하나의 텍스트 파일이 생성되었다. 따라서, 총 682개(1학년 216개, 2학년 261개, 3학년 205개)의 텍스트 파일들로 구성된 과학교과서 코퍼스가 구축되었다.

본문 텍스트만을 추출한 이유는 교과서에 여러 연습 활동 및 추가 학습 활동들이 포함되어 있지만, 기본적인 교수 학습 방향을 결정하는 가장 핵심적인 텍스트가 본문이기 때문이다(Kim, Ryu, & Kang, 2005; Ryu & Jeon, 2020). 또한, 중학교 교과서를 선정한 이유는 다음과 같다. 먼저, 초등에서 고등에 이르기까지 전 학년 학습자들 모두 타교과목에 비하여 과학교과서를 이해하는 데 여러 어려움을 겪고 있다. 하지만, 그중에서도 과학을 접하는 초기 단계인 초등 및 중등 수준에서 학습자들은 인지 및 언어 발달이 급격하게 진행되고 배경 지식 또한 상대적으로 불충분하기에 교육학적 관점에서 텍스트 자체 수준(텍스트 특성)에서의 개입(intervention)이 보다 중요해지는 시기이다(Best et al., 2005). 초등학교 과학교과서를 대상으로 이루어진 연구는 어느 정도 이루어졌지만(e.g., Kang & Park, 2009), 중학교 교과서를 대상으로 언어적 특성을 분석한 연구는 거의 찾아보기 어렵다. 또한, 초등학교 교과서에 비해 중학교 교과서 텍스트의 학습 내용이 양적으로도 훨씬 더 풍부하기 때문에 언어적 특성에 따른 효과가 보다 클 것이라 기대된다. 이러한 이유에서 본 연구에서는 중학교 과학교과서를 분석 대상으로 선정하였다.

3. 분석 방법

본 연구에서는 중학교 1, 2, 3학년 과학교과서에 수록된 본문 텍스트의 특성을 분석하기 위해서 세 개 집단에 대한 교과서 코퍼스 파일을 구축하였다. 이렇게 구축된 텍스트 파일들을 한국어 분석 시스템인 Auto-Kohesion 시스템을 통해 분석하였다. 전술한 것처럼, 본 연구

에서는 Auto-Kohesion 시스템이 제공하는 측정치 중에서 에세이 연구에 사용되는 텍스트 구조 분석 측정치를 제외한 총 20개의 측정치를 선택해서 통계분석에 이용하였다. 구체적으로 본 연구에서는 Auto-Kohesion 각각의 측정치에 대한 세 개 집단의 언어적 특성 차이를 분석하기 위해서 집단의 분산(variance) 차이에 기반한 집단 간의 평균 차이 검증에 이용되는 분산분석(analysis of variance, ANOVA) 통계방법을 이용하였다. 분산분석 시, 독립변인은 학년 수준(중학교 1, 2, 3학년)이고 종속변인은 Auto-Kohesion 각각의 측정치였다. Auto-Kohesion 측정치에 대한 설명 및 산출 방식은 아래 해당 측정치 분석결과 부분에 제시하였다. 본 연구에서는 SPSS 21.0을 활용해 통계분석을 수행하였다. 먼저 각 집단의 차이 검증을 위해 일원분산분석(one-way ANOVA)을 실시하였다. 일원분산분석 수행 후 학년 간에 유의미한 차이가 발견될 경우, 구체적으로 어느 학년 간의 차이가 통계적으로 유의미하게 나타났는지를 추가적으로 검증하기 위해 Bonferroni 사후 분석(post hoc analysis)을 실시하였다. 모든 분석에 대해 유의수준(significance level)은 5%($p = .05$)로 설정되었다.

III. 분석결과

중학교 1~3학년 과학교과서 텍스트 간의 언어적 특성을 Auto-Kohesion 시스템이 제공하는 20개의 측정치를 기반으로 통계분석을 수행한 결과를 제시하면 다음과 같다.

1. 텍스트 표층 구조 측정치

텍스트 표층 구조 측정치에는 어절 수, 문장 수, 평균 어절 길이 및 평균 문장 길이가 포함된다. 평균 어절 길이는 각 어휘에 대한 어절의 수를 전체 어휘수로 나눈 평균 값으로 산출된다. 평균 문장 길이는 텍스트 속에 포함된 전체 어휘수를 전체 문장수로 나눈 평균 값으로 산출된다. 이와 같은 측정치들은 선행 연구에서 교과서의 이독성(readability)을 산출하기 위해 널리 활용되었다(Ryu & Jeon, 2020). 텍스트 표층 구조 측정치에 대해서 중학교 과학교과서 텍스트를 분석한 결과 아래 <표 1>과 같이 나타났다. 표 안의 수치는 각 학년에 포함된 각각의 텍스트 파일(각각의 단락)에 대한 해당 측정치 값들의 평균(mean) 점수이고 괄호 안의 수치는 표준 편차(standard deviation)이다.

<표 1>에 나타난 것처럼, 각 단락에 대한 평균 어절 수와 문장 수는 학년 간 유의미한 차이가 없었다(all $ps > .05$). 이러한 결과는 학년이 올라갈수록 더 많은 내용을 배우고 그에 따라 단락의 길이가 길어질 것이라는 일반적 기대와는 상충된다. 1학년에 비하여 2학년 교과서에서는 하나의 단락으로 구성된 텍스트 파일이 45개나 증가하

Table 1. Results on basic counts

측정치	M1(n=216)	M2(n=261)	M3(n=205)	F	p	posthoc
어절 수	173(80.6)	165(78.8)	177(78.5)	1.322	.267	N/A
문장 수	13.3(6.35)	12.8(6.44)	13.1(5.80)	.416	.660	N/A
평균 어절 길이	2.77(.15)	2.80(.14)	2.80(.14)	2.857	.058	N/A
평균 문장 길이	13.2(2.03)	13.2(1.97)	13.6(1.87)	3.07	.047*	N/A

* $p < .05$

여 다루는 내용이 더 많아졌을 것이라 사료되지만 3학년 교과서의 경우 오히려 1학년 교과서보다도 더 적은 결과가 나타났다.

평균 어절 길이는 1학년 교과서에 비해 2, 3학년 교과서에서 더 길어지는 경향성을 보였지만($p = .058$), Bonferroni 사후 분석 결과, 통계적으로 유의미한 차이를 나타내는 두 개 집단 비교 짝(pair)은 발견되지 않았다. 마찬가지로, 평균 문장 길이 또한 1, 2학년 교과서에 비하여 3학년 교과서에서 조금 더 길어지는 경향성을 보였지만($p < .05$), 사후 분석 결과 통계적으로 유의미한 차이를 나타내는 비교 짝은 없었다. 위 결과는 텍스트 표층 구조 측정치에 대해서는 학년이 증가함에 따라 눈에 띄는 변화가 발견되지 않는다는 것을 보여준다.

2. 어휘 정보 측정치

본 연구의 분석에 활용된 Auto-Kohesion 시스템에는 국립국어원의 21세기 세종계획에 따라 구축된 세종코퍼스를 기반으로 형성된 각 어휘에 대한 빈도를 담고 있는 데이터베이스가 포함되어 있다. Auto-Kohesion 시스템은 이를 활용하여 여러 어휘 빈도 정보 측정치를 계산한다. 구체적으로, 세종코퍼스에 기반한 구축된 어휘빈도 데이터베이스는 세종코퍼스 내에 포함된 전체 어휘들에 대한 빈도정보로 구성된다. Auto-Kohesion 시스템은 이 데이터베이스에 있는 어휘 빈도수를 활용하여 분석하고자 하는 텍스트에 포함된 어휘들에 대한 빈도 값을 산출한다. 부연하면, 어휘빈도 측정치는 분석대상 텍스트 내에 포함된 어휘들에 대응되는 데이터베이스 내의 어휘들에 대한 빈도수의 평균 값을 지칭한다(Ryu & Jeon, 2020). 어휘 빈도 측정치에는 전체 단어 및 내용에 대한 빈도 값과 그 값에 자연 로그를 취한 값이 포함된다. 원빈도와 자연 로그를 취한 값 모두를 포함하는 이유는 원점수를 로그 값으로 변환하면 극단값들이 보정되면서 전체적인 값들의 분포가 정규분포에 더욱 근접하게 되면서 통계 분석을 위한 가정(assumption)에 잘 충족되기 때문이다. 어휘 빈도 값은 수치가 높을수록 상대적으로 고빈도(high-frequency) 단어이다. 일반적으로, 텍스트 내에 고빈도 어휘들이 많이 존재하면 텍스트의 이해도는 증가하고 난이도는 감소하기 때문에 저빈도 어휘들이 많이 사용된 텍스트에 비해서 정합성 형성이 상대적으로 잘 일어날 수 있다(Graesser, Jeon, Yan, & Cai, 2007). 어휘 다양성 측정치로는 타입-토큰 비율(type-token ratio) 값이 제공된다. 타입-토큰 비율에서 타입이란 텍스트 파일에 처음으로 등장한 개개 단어(unique word)의 발생 빈도수(따라서 타입 값은 항상 1임)를 가리키며, 토큰이란 텍스트 파

일에 포함된 모든 개별 단어(individual word)가 텍스트 내에 반복되어 사용된 빈도수를 가리킨다(Ryu & Jeon, 2020). 따라서, 어휘가 반복적으로 사용되면 타입의 수는 변하지 않지만 토큰의 수는 증가하기 때문에 타입-토큰 비율 값이 작아지게 된다. 반면 어휘가 반복되지 않고 한 번씩만 사용될 경우 타입-토큰 비율값은 1에 가깝게 된다. 즉, 타입-토큰 비율은 본문에 얼마나 다양한 단어가 사용되었는지를 나타내는 값으로, 값이 높을수록 어휘 다양성 정도가 증가한다. 텍스트 내에 다양한 어휘들이 많이 사용될수록 이러한 어휘들에 대한 처리 부담이 증가하기 때문에 정합성 형성이 상대적으로 잘 일어나지 않거나 방해받을 수 있다(Graesser, Jeon, McNamara, & Cai, 2008; Graesser, McNamara, Louwerse, & Cai, 2004). 이처럼 어휘 빈도와 어휘 다양성과 같은 어휘 정보 측정치는 교과서 텍스트의 언어적 특성을 비교 분석하는 선행 연구에서 활용되는 측정치로서 텍스트 이해에 직간접적으로 상당한 영향을 미친다고 알려져 있다(Graesser, Jeon, Yan, & Cai, 2007; Ryu & Jeon, 2020). 전술한 어휘 정보 측정치에 대해 중학교 1~3학년 과학교과서를 비교한 결과는 <표 2>에 제시되어 있다.

어휘 빈도 측정치 중 전체 단어에 대한 원빈도 값과 로그 값을 보면 학년이 증가함에 따라 값이 낮아지는 경향성이 나타났다($p = .073$ for the raw score, $p = .060$ for the logarithmic score). 즉, 학년이 증가함에 따라 다소 빈도가 낮은 어려운 단어가 더 많이 사용된 것을 알 수 있다. 반면, 내용에 대한 분석 결과는 학년 간에 통계적으로 유의미한 차이가 발견되지 않았다($ps > .05$). 학년이 증가함에 따라 학습자의 인지 능력 및 언어 능력이 발달되고, 또한 학습자가 더욱 전문적으로 과학 내용에 대해 배우기 때문에 교과서 내에 보다 빈도가 낮은 어휘가 점차적으로 더 많이 사용되는 경향이 어느 정도 반영된 것으로 이해할 수 있다. 그러나 추후 교과서 개발 시에 내용에 대한 빈도 또한 고려되어 교과서가 집필된다면 학습자의 학술적 어휘 발달에 기여할 것이라 기대된다.

어휘 다양성을 나타내는 타입-토큰 비율 분석 결과, 학년 간에 통계적으로 유의미한 차이가 나타났다($p < .05$), Bonferroni 사후 검증 결과 차이가 나는 지점은 1학년과 2학년 교과서 간 차이값이었다($p < .05$). 즉, 1학년 교과서는 2학년 교과서에 비해 보다 다양한 어휘를 포함하고 있는 것을 알 수 있다. 이러한 결과는 저학년 교과서에 반복적으로 같은 어휘가 사용되어 학습자들의 부담이 낮아지고 고학년이 되면서 보다 여러 주제가 다루어지면서 다양한 어휘가 사용될 것이라는 일반적 기대와는 상충된다.

Table 2. Results on lexical features

측정치	M1(n=216)	M2(n=261)	M3(n=205)	F	p	posthoc
어휘 빈도						
전체 단어(원빈도)	4626(1893)	4625(2147)	4240(1942)	2.632	.073	
전체 단어(자연로그)	8.36(.42)	8.34(.47)	8.26(.44)	2.832	.060	
내용어(원빈도)	2615(1003)	2464(985)	2516(1231)	1.190	.305	
내용어(자연로그)	7.80(.38)	7.73(.41)	7.73(.44)	1.991	.137	
어휘 다양성						
타입-토큰 비율	.63(.11)	.61(.10)	.61(.10)	3.542	.029*	1>2

* $p < .05$

3. 통사적 복잡성 측정치

통사적 복잡성 측정치는 수식언 비율과 문장 성분 비율 값을 포함한다. 먼저, Auto-Kohesion 시스템이 제공하는 수식언 비율은 문장 내 제시된 체언 및 용언을 수식하는 관형사 및 부사의 평균 비율 값을 통해 계산된다(Ryu & Jeon, 2020). 수식언 비율 값이 높을수록 수식하는 관형사 및 부사가 많아지는 것을 뜻한다. 이는 한 문장에 보다 많은 정보가 압축되어 있고, 이로 인해 작업 기억(working memory)의 부담이 증가하므로 처리하기 복잡한 문장이 된다. 따라서, 이 값이 큰 경우 텍스트의 정합성이 상대적으로 낮게 형성될 수 있다(Graesser, Jeon, McNamara, & Cai, 2008; Ryu & Jeon, 2020). 문장 성분 비율 값은 문장의 통사 트리 구조(tree structure)를 구성하는 문장 성분(constituent)의 수를 해당 문장에 포함된 전체 어휘의 수로 나눈 값이다(Graesser, McNamara, Louwerse, & Cai, 2004; Ryu & Jeon, 2020). 따라서 문장 성분 비율 값이 증가할수록 문장 성분이 더욱 많아지므로 더욱 복잡한 문장이 된다. 요약하면, 이러한 통사적 복잡성 값의 증가는 처리부담을 가중시켜 결과적으로 텍스트의 정합성을 방해하거나 지연시키는 결과를 초래한다(Graesser, McNamara, Louwerse, & Cai, 2004). 교과서 개발 및 집필에 대한 일반적 기대는 학습자들의 인지 및 언어 발달에 맞추어 점점 더 복잡한 문장을 학습자들에게 제시하여 추후 고등교육에 대비할 수 있는 능력이 발달되도록 하는 것이다. 이에 따르면, 학년이 증가할수록 수식언 비율 및 문장 성분 비율값으로 측정된 통사적 복잡도가 증가하여야 한다. 중학교 1~3학년 과학교과서에 수록된 본문 텍스트의 통사적 복잡성을 비교 분석한 결과는 <표 3>에 제시되어 있다.

수식언 비율과 문장 성분 비율 값에 대한 분석 결과를 보면, 두 측정치 모두 학년 간에 통계적으로 유의미한 차이가 발견되었는데($ps < .05$), 학년이 증가함에 따라 통사적 복잡성이 전반적으로 감소하는 것을 확인할 수 있다. 구체적으로 어떤 학년 간에 유의미한 차이가 발견되었는지 추가적으로 검증하기 위하여 Bonferroni 사후 검증을 실시하였다. 그 결과, 수식언 비율에 대해서는 1학년 교과서가 2, 3학년 교과서보다 통사적으로 더욱 복잡한 문장을 더 많이 포함하였고($p < .05$), 문장 성분 비율에 대해서는 1, 2학년 교과서가 3학년 교과서에 비하여 통사적으로 복잡한 문장을 많이 사용한 것으로 나타났다($p < .05$). 이러한 결과는 앞서 전술한 교과서 개발 및 편찬에 대한 일반적인 기대와는 상충하는 결과이다. 따라서 이러한 결과가 추후 교과서 개발 시 고려되기를 기대한다.

Table 3. Results on syntactic complexity

측정치	M1(n=216)	M2(n=261)	M3(n=205)	F	p	posthoc
수식언 비율	.05(.03)	.04(.02)	.04(.02)	9.178	.000*	1>2,3
문장 성분 비율	1.73(.08)	1.75(.07)	1.68(.10)	36.547	.000*	1,2>3

* $p < .05$

Table 4. Results on noun overlap

측정치	M1(n=216)	M2(n=261)	M3(n=205)	F	p	posthoc
명사 반복 (인접 문장)	.85(.15)	.88(.14)	.90(.12)	5.353	.005*	1<3
명사 반복 (전체 문장)	.73(.17)	.72(.17)	.76(.16)	3.414	.033*	2<3

* $p < .05$

4. 명사 반복 측정치

텍스트 정합성은 명사 반복 측정치와 연결어 측정치, 그리고 대명사 측정치를 통해 조절될 수 있다(Ryu & Jeon, 2020). 명사 반복 측정치는 문장 간에 명사가 공유되는지를 나타낸 값으로, 값이 증가할수록 명사가 반복되는 문장 쌍(sentence pair)이 많다는 것을 가리킨다. 문장 간 명사가 반복되어 사용되면 그렇지 않은 경우에 비해 독자의 처리 부담이 낮아지게 된다. 왜냐하면 두 문장이 서로 같은 대상을 통해 연결되어 있기 때문에 독자가 두 문장의 관계를 추론할 필요성이 감소하고, 이로 인해 작업 기억의 부담도 감소하기 때문이다. 이처럼 명사 반복 측정치 값이 증가할수록 보다 짜임새 있는 텍스트가 되기 때문에 텍스트 정합성이 높아지게 된다(Cirilo, 1981; Haviland & Clark, 1974; Graesser, Jeon, Yan, & Cai, 2007). 앞서 논의한 것처럼, 텍스트 정합성은 학습자들이 과학교과서를 이해하는 데 큰 영향을 미칠 수 있는 핵심적인 텍스트 요인이다. 인접 문장 및 전체 문장에 대한 문장 간 명사 반복에 대해 중학교 1~3학년 과학교과서 본문 텍스트를 분석한 결과는 <표 4>에 제시되어 있다.

Auto-Kohesion 시스템이 제공하는 인접 문장 쌍에 대한 명사 반복 측정치는 명사 반복이 된 인접 문장 쌍들의 수를 전체 인접 문장 쌍들의 수로 나눈 값을 지칭한다. 반면에, 전체 문장 쌍에 대한 명사 반복 측정치는 명사 반복이 된 문장 쌍들의 수를 전체 문장 쌍들의 수로 나눈 값을 나타낸다(Ryu & Jeon, 2020). 인접 문장 및 전체 문장에 대한 명사 반복 측정치 분석 결과, 저학년 교과서에 비해 고학년 교과서에서 상대적으로 더 많은 문장 쌍에서 명사가 반복되어 사용된 것으로 나타났다. 다시 말하면, 고학년 교과서 텍스트에서 오히려 더 짜임새 있고 정합성이 높은 문장들이 사용되었다($ps < .05$). 구체적으로 인접 문장 간 명사 반복 값에 대해서는 1학년 교과서보다 3학년 교과서에서 정합성이 높은 텍스트가 사용되었다($p < .05$). 2학년 교과서와 1학년과 3학년 교과서 간에는 통계적으로 유의미한 차이가 발견되지 않았다($ps > .05$). 이러한 결과는 저학년 텍스트에서는 문장 간의 관계가 언어적 장치들을 통해 연결되는 정도가 상대적으로 낮고, 고학년 텍스트에서는 문장 간의 관계가 언어적 장치로 인해 연결되는 정도가 높다는 것을 나타낸다. 이러한 결과는 과학교과서 텍스트의 언어적 특성이 학습자의 특성을 고려하여 단계적으로 조절되어야 한다는 원칙(Vygotsky, 1978)을 따르고 있지 않은 것으로 사료된다. 이러한 결과는 학습자들의 상대적으로 낮은 인지 및 언어 발달 수준, 그리고 과학에 입문하면서 갖게 되는 낮은 배경지식을 보완하기 위해

저학년 교과서에 상대적으로 정합성이 높은 텍스트가 제시될 필요성을 시사한다. 다시 말하면, 학습자의 발달수준에 맞추어 점차적으로 정합성이 낮은 텍스트를 제시해주면서 학습을 촉진시키는 것이 바람직함을 시사한다.

5. 연결어 측정치

연결어 측정치는 전체 접속부사, 인과적 접속부사 및 부가적 접속부사를 포함한다. 접속부사는 문장 내 혹은 문장 간 두 요소 사이의 관계를 명시적으로 제시해주는 언어 장치이다(Ryu & Jeon, 2020). 두 개별적인 요소가 서로 언어적으로 연결되지 않고 제시되는 경우에 비하여 두 요소가 접속부사를 통해 명시적으로 연결되면 독자는 두 요소 간의 논리적 관계를 파악하는데 도움을 받을 수 있다. 특히, 인과적 접속부사는 문장 간의 인과적 연결을 촉진시키기 때문에 텍스트의 정합성을 증가시키는 데 기여한다(Caron, Micko, & Thuring, 1988; Millis & Just, 1994). 이처럼, 접속부사의 사용은 텍스트 내 요소를 짜임새 있게 엮어주므로 접속부사가 사용되면 텍스트 정합성이 증가한다(Ryu & Jeon, 2020). 연결어 측정치에 대해 중학교 1~3학년 과학교과서를 분석한 결과가 <표 5>에 제시되어 있다. Auto-Kohesion 시스템이 제공하는 접속부사 측정치는 텍스트의 길이에 의한 효과를 통제하기 위해 어휘 100개 당 텍스트에 나타나는 수로 계산된다. 이러한 점수는 발생점수(incidence score)라 지칭되며 텍스트의 길이를 통제하기 위한 대안으로 많이 사용된다(Graesser, McNamara, Louwerse, & Cai, 2004). 표 5에 제시된 점수는 각 접속부사에 대한 발생점수를 나타낸다.

분석 결과, 전체 접속부사와 인과적 접속부사에 대해서는 학년 간에 통계적으로 유의미한 차이가 발견되지 않았다($ps > .05$). 인과적 접속부사는 텍스트 내 두 요소 간의 관계가 인과적으로 연결되어 있다는 정보를 제시함으로써 학습 내용에 대해 부족한 독자의 배경지식 역할을 보완해줄 수 있다. 이 때문에, 과학을 입문하는 단계에 있는 상대적으로 배경지식이 부족한 저학년 학습자들을 위해 저학년 교과서에 인과적 접속부사가 많이 사용될 필요성이 기대된다. 하지만, 이러한 기대와 다르게 학년 간에 사용된 인과적 접속부사의 비율에 대해 주목할 만한 차이가 발견되지 않았다.

반면, 부가적 접속부사는 2학년 교과서보다 1학년 교과서에 더 많이 사용된 것으로 나타났다($p < .05$). 부가적 접속부사는 인과적 접속부사에 비하여 두 요소가 어떤 관계로 연결되는지에 대해 상대적으로 중립적인 정보를 제공한다. 이에 따라, 인과적 접속부사에 비해 텍스트 정합성에 미치는 영향이 상대적으로 낮다(Millis & Just, 1994). 일반적으로 부가적 접속부사가 많이 사용되면 복문 수의 증가로 인해 텍스트의 통사적 복잡성이 증가할 수 있다. 일반적인 기대와 달리, 1학년 교과서에 부가적 접속부사가 많이 사용된 이유는 저학년일수록 과학 원리를 이해하는데 구체적인 사례(이야기 형식) 등이 필요하므로 이러한 사례 등을 기술하는 데 상대적으로 복잡한 문장 등이 사용되었기 때문이라고 할 수 있겠다.

6. 대명사 측정치

대명사 측정치에는 전체 대명사, 1인칭 대명사, 2인칭 대명사, 그리고 3인칭 대명사 측정치가 포함된다. 일반적으로 텍스트 내에 대명사가 많이 사용되면, 대명사가 가리키는 지시대상을 성공적으로 해결(pronoun resolution)해야만 두 문장 간의 연결이 적절히 이루어지기 때문에 많은 양의 대명사 처리로 인해 텍스트의 정합성이 낮아지는 것으로 이해할 수 있다(Graesser, Jeon, Yan, & Cai, 2007; Ryu & Jeon, 2020). 이러한 경향은 특히 3인칭 대명사의 경우에 그러하다. 1, 2인칭 대명사의 경우는 지시대상이 한정적이기 때문에 바로 처리할 수 있지만, 3인칭 대명사의 경우는 1, 2인칭 대명사에 비해 상대적으로 지시대상이 많기 때문에 이로 인해 텍스트 정합성이 낮아질 수 있다. 이에 따라, 대명사의 종류에 따라서 상반된 예상이 가능한데, 1, 2인칭 대명사의 경우에는 저학년 교과서에 더욱 많이 분포되어 있고, 3인칭 대명사의 경우에는 고학년 텍스트에 보다 많이 사용될 것으로 기대된다. 대명사 측정치에 대해 중학교 1~3학년 과학교과서를 분석한 결과가 <표 6>에 제시되어 있다.

분석 결과를 보면, 전체 대명사 및 1인칭 대명사는 학년 간에 통계적으로 거의 유의미한 차이(marginally significant)가 나는 경향성을 보인다. Bonferroni 사후 검증 결과, 전체 대명사 및 1인칭 대명사가 3학년 교과서보다 1학년 교과서에 더 많이 사용되었고, 2학년 교과서보다 3학년 교과서에 많이 나타났다($ps < .05$). 2인칭 대명사 및 3인칭

Table 5. Results on Connectives

측정치	M1(n=216)	M2(n=261)	M3(n=205)	F	p	posthoc
전체 접속부사	.78(.77)	.65(.74)	.77(.79)	2.335	.098	N/A
인과적 접속부사	.21(.39)	.22(.44)	.28(.43)	1.712	.181	N/A
부가적 접속부사	.58(.67)	.42(.63)	.49(.64)	3.294	.038*	1>2

* $p < .05$

Table 6. Results on Pronouns

측정치	M1(n=216)	M2(n=261)	M3(n=205)	F	p	posthoc
전체 대명사	.27(.62)	.17(.44)	.21(.41)	2.701	.068	1>3>2
1인칭 대명사	.27(.62)	.16(.43)	.20(.39)	2.821	.060	1>3>2
2인칭 대명사	.00	.00	.00	.	.	N/A
3인칭 대명사	.00(.03)	.00(.05)	.01(.10)	1.844	.159	N/A

* $p < .05$

대명사는 전 학년에 걸쳐 거의 사용되지 않았으므로 전체 대명사의 사용은 1인칭 대명사의 사용을 반영한 결과로 이해된다. 1인칭 대명사가 많이 사용되면 학술적(academic)이기보다 일상적(everyday) 글에 가깝기 때문에, 1인칭 대명사의 사용은 아직까지 과학이라는 학술적 언어에 익숙하지 않은 저학년 학습자들에게 더욱 적합하다. 따라서 3학년 교과서에서의 1인칭 대명사 사용의 증가는 고등 교육에 대비해 점점 학술적인 글을 이해하는 능력이 개발되어야 하는 것을 기대하게 하는 3학년 교과서에는 적절하지 않을 수 있다는 것을 시사한다.

2인칭 대명사는 전 학년 교과서에 한 번도 사용되지 않았기 때문에 통계 분석이 수행되지 않았다. 3인칭 대명사 또한 전 학년 교과서에 거의 사용되지 않았으며 학년 간에 통계적으로 유의미한 차이도 발견되지 않았다($p > .05$). 3인칭 대명사가 많이 사용되는 경우 학습자들은 1, 2인칭 대명사에 비해 상대적으로 대명사가 지칭하는 대상을 찾기 위해 더 많은 인지 처리와 노력을 필요로 한다. 이 과정에서 대명사 처리로 인한 추론 능력 등이 향상될 것으로 기대된다. 따라서, 추후 교과서 개발 시에 이러한 점이 반영되어 대명사로 인한 교과서 텍스트의 정합성이 학습자들의 인지 발달 수준에 맞게 좀더 체계적으로 조절될 수 있도록 교과서를 집필할 필요가 있다.

IV. 결론 및 제언

텍스트 이해 및 텍스트를 통한 학습 과정을 설명하는 구성주의 접근에서는 독자와 텍스트 특성의 상호작용을 강조한다(Graesser & Clark, 1985; van Dijk & Kintsch, 1983). 이에 따라 학습자들의 이해 및 학습을 촉진시키기 위해서는 교과서 개발 시에 독자의 특성을 고려하여 텍스트 특성을 조절하는 것이 중요하다. 현재 연구에서는 과학교과서 텍스트에 이러한 원칙이 잘 반영되어 있는지를 분석하기 위해, 중학교 1~3학년 과학교과서의 본문 텍스트를 추출한 후, Auto-Kohesion 시스템이 제공하는 20개의 언어적 측정치를 통해 학년 간 언어적 특성 차이를 비교 분석하였다. 분석에 활용된 측정치는 어절 및 문장 수, 평균 어절 길이 및 평균 문장 길이를 측정하는 텍스트 표층 구조 측정치, 내용어 및 전체 단어에 대한 어휘 빈도(원빈도 및 자연로그를 취한 빈도)와 타입-토큰 비율로 추산된 어휘 다양성을 포함하는 어휘 정보 측정치, 수식언 비율과 문장 성분 비율로 계산되는 통사적 복잡성 점수를 포함한다. 또한, 정합성 관련 측정치로써 명사 반복(인접 문장 및 전체 문장 간) 측정치, 전체 접속부사, 인과적 및 부가적 접속부사를 측정하는 연결어 측정치, 그리고 전체 대명사, 1인칭, 2인칭 및 3인칭 대명사의 수를 산출하는 대명사 측정치를 포함한다.

먼저 텍스트 표층 구조, 어휘 정보 및 통사 측정치에 대한 결과를 요약하면, 평균 어절 및 문장 길이는 학년이 올라갈수록 증가하는 경향을 보였고, 전체 단어에 대한 원빈도 및 로그빈도는 낮아지는 경향성을 보였다. 이러한 결과는 학년 증가에 따른 학습자들의 인지 능력과 언어 능력의 발달 및 보다 전문적인 과학 관련 단어를 학습할 필요성에 의한 교과서 텍스트의 언어적 변화로 이해할 수 있다. 반면, 한 단락에 실린 어절 수와 문장 수의 평균 값에는 학년 간에 통계적으로 유의미한 차이가 없었고, 내용어에 대한 어휘빈도 값에도 유의미한 차이가 발견되지 않았다. 타입-토큰 비율에서는 1학년 교과서에서 2, 3학년 교과서보다 더 다양한 어휘가 사용된 것으로 나타났다. 또한,

통사적 복잡성을 측정하는 수식언 비율 및 문장 성분 비율에서는 전 반적으로 저학년 교과서에 고학년 교과서보다 통사적으로 복잡한 문장이 사용되었다. 이러한 측정치 결과에 기초하여 과학교과서의 구성과 과학교수학습에 대한 함의점을 간략히 제시하면 다음과 같다. 과학교과서 텍스트를 개발할 때 예상 독자층의 개인 특성을 고려하며 텍스트 내에 포함된 다양한 언어적 특성을 체계적으로 조절 및 제시하는 것이 중요하다. 가령, 저학년 학습자들은 고학년에 비해 상대적으로 인지 및 언어 능력이 미숙하고 과학에 입문하는 단계에 있기 때문에 배경지식도 부족하다. 이에 따라, 텍스트 표층 구조, 어휘 및 통사적 특성에 대해서 상대적으로 처리하기 쉬운 어휘들(저빈도 어휘들) 및 언어적 통사 구조가 저학년 교과서에 배치되고 점차적으로 학년이 올라갈수록 어려운 구조가 제시되는 것이 학습 효과를 최대화하는 데 유용할 것이다(Vygotsky, 1978). 과학교과서 교수학습의 목적이 국어교과서처럼 언어교육 학습에 있는 것은 아니지만, 본 연구의 결과가 시사하는 것처럼 텍스트의 언어적 속성이 교과서 이해에 미치는 영향이 크기 때문에 학년에 따른 교과 내용 배치뿐만 아니라 언어적 특성에 기반해 텍스트의 전체적인 이해도 및 난이도를 통제하는 것이 중요하다고 하겠다. 그러나 현재 연구 결과에서는 교과서 개발에서 기대되는 이러한 원칙이 잘 반영되지 않은 것으로 나타났다. 구체적으로, 학년 증가에 따른 저빈도 단어 사용의 증가(내용어), 어휘 다양성 증가, 통사적 복잡성 증가 등이 고려되어 교과서가 집필되기를 기대한다.

다음으로, 정합성 관련 측정치에 대한 분석 결과를 요약하자면 다음과 같다. 먼저, 명사 반복 측정치에 대해서는 1학년 교과서보다 3학년 교과서에서 더 많은 문장 간에 명사가 반복적으로 사용되었다. 그리고, 전체 문장에 대해서는 2학년 교과서보다 3학년 교과서에 더 많은 문장 간에 명사 반복이 이루어졌다. 다시 말하면, 저학년 교과서에 오히려 정합성이 낮은 텍스트가 제시되고 고학년 교과서에 정합성이 높은 텍스트가 제시되었다. 그 다음으로, 전체 접속부사와 인과 접속부사의 사용에서는 학년 간에 유의미한 차이가 없었고, 부가적 접속부사는 2학년 교과서보다 1학년 교과서에 더욱 많이 사용되었다. 마지막으로 대명사 측정치 분석 결과, 전체 대명사 및 1인칭 대명사의 사용은 1학년 교과서에 가장 두드러지게 나타났고, 그 다음으로는 3학년, 2학년 순이었다. 접속사는 텍스트 내 두 문장 간의 연결 관계를 명시적으로 나타내 주기 때문에, 접속사가 많이 사용되면 텍스트 정합성이 올라간다. 반면, 3인칭 대명사의 사용이 많아지게 되면 독자가 대명사의 지시대상(referent)을 추론해야 하므로 텍스트 정합성이 낮아진다. 이를 고려하면 인과 접속부사 및 3인칭 대명사에 대해, 저학년 교과서에 고학년 교과서에서나 등장할 만한 정도의 정합성이 낮은 텍스트가 사용되었음이 시사된다. 이러한 결과를 바탕으로 과학교과서 개발에 대해 제시하면 다음과 같다. 일반적으로, 정합성이 낮은 텍스트는 독자에게 배경지식을 활용하여 추론(inference)하도록 유도한다(Graesser, Jeon, Yan, & Cai, 2007). 이 때문에 이러한 측정치에 대한 현재의 1학년 과학교과서의 내용은 인지 및 언어 능력이 상대적으로 낮고 배경 지식도 부족한 저학년 학습자들에게 적합하지 않다고 판단된다. 따라서, 추후 교과서 집필 시 이러한 내용이 반영되어 학년 간 교과서의 난이도가 체계적으로 조절되기를 기대한다. 특히, 명사 반복과 같은 측정치는 정합성 형성 및 텍스트 이해에 큰 영향을 주는 요소임에도 불구하고 현실적으로 텍스트 이해자가 텍스트 이해 시

이러한 요소들을 의식하면서 이해를 하기는 불가능하다. 이 점에서 본 연구의 결과가 주는 시사점이 중요하다고 볼 수 있겠다. 물론 점인정 체계에 따라서 교과서를 개발해야 하는 한국의 상황을 고려할 때, 텍스트 분량 제한으로 인해 본 연구에서 분석한 다양한 언어 측정치를 사용하는 데 한계점이 있었지만 차후 교과서 개발 시, 교과서 개발자들이 이러한 명사 반복 측정치의 중요성을 인식하여 교과서 내용 구성 등 여러 가지 고려할 부분이 많겠지만 이러한 측정치를 반영하여 교과서를 집필하기를 기대한다. 또한 이러한 결과를 교수학습 측면에서 고려해보면, 과학 수업 시에 과학 교사는 이러한 연구 결과를 인식하여, 중학교 저학년 수업 시 상대적으로 낮은 정합성으로 인한 이해 난이도를 감소시키기 위해 좀 더 쉬운 방법으로 교과서 내용을 설명할 필요가 있겠다. 반면에, 고학년 수업 시간에는 교과서의 정합성이 다소 높기 때문에 좀 더 어려운 과학 내용을 학습시킬 수 있겠다. 그리고 이러한 결과를 중학교 1~3학년 국어교과서의 언어적 특성을 비교 분석한 Ryu와 Jeon(2020)의 연구 결과를 고려하면 또 다른 시사점이 도출된다. 이 연구에서 중학교 1, 2, 3학년 국어교과서 텍스트의 인접 문장 간 명사 반복 점수 평균은 .45이고, 전체 문장 간 명사 반복 점수 평균은 .29였다. 이에 비하면, 현재 연구에서 분석한 과학교과서의 인접 문장 간 명사 반복 점수 평균은 .88, 전체 문장 간 명사 반복 점수 평균은 .74로 국어교과서에 비해 상당히 높은 수치이다. 일반적으로, 교과목의 목적에 따른 교과서 텍스트의 특성이 상이하다. 다시 말하면 국어과의 경우에는 국어 능력의 신장을 주된 목표로 삼는 반면, 과학과의 경우에는 새로운 정보의 습득을 주된 목표로 삼고 있기 때문에 과학교과서의 텍스트 정합성이 국어교과서에 비해 높다 (Lee, 2011; McNamara *et al.*, 2009; Ozuru *et al.*, 2005; Staples, Egbert, Biber, & Gray, 2016). 이를 고려하더라도, 중학교 과학교과서의 명사 반복 수치는 상대적으로 높다고 판단된다. 따라서, 한편으로 과학교과서의 정합성을 보정하기 위해 저학년 텍스트의 정합성을 높이는 시도보다 고학년 텍스트의 정합성을 낮추는 방향이 더욱 바람직할 것이라 기대된다.

이러한 결과를 정리하면, 대체로 어절 및 문장 길이, 어휘 빈도와 같은 피상적으로 두드러지는 특성에 대해서는 단계적인 조절이 이루어졌지만, 그 외의 많은 언어적 특질에 대해서는 학년이 증가함에 따라 체계적인 조절이 되지 않은 것으로 나타났다. 이러한 결과 또한 차후 교과서 개발에 대한 시사점을 제공한다. 특히 정합성 형성에 상대적으로 큰 영향을 주는 여러 정합성 측정치들(명사반복, 대명사, 접속사)이 교과서 개발 과정에서 충분히 고려되지 않은 것으로 시사된다. 이러한 이유 중 하나는 교과서 텍스트를 양적으로 분석한 선행 연구들이 대체로 어절 및 문장 길이, 어휘 빈도 수준에서의 분석에 치중되어 있었기 때문이다. 이는 텍스트 이해에 영향을 미치는 요인에 대한 이해가 부족하거나, 이해가 있다고 하더라도 이러한 피상적인 측정치를 넘어 더욱 심층 수준에서의 측정치에 기반하여 텍스트 특성을 측정하는 것의 어려움을 반영하는 것이기도 하다(Suh & Ryu, 2014). 또한, 이러한 한계가 교과서 개발 및 평가 시에도 작용하여 주로 이러한 피상적 측정치를 통해 교과서 개발 및 평가가 이루어졌을 가능성이 있다. 그러나, 교과서를 통한 학습의 효율을 최대화하기 위해서는 정합성과 같은 심층 수준에서의 측정치를 포함하여 다각도로 텍스트의 특성을 체계적으로 조절하는 것이 중요하다. 따라서 차후 교과서 집필 시 본 연구의 결과가 반영되고 좀 더 체계적으로 과학

교과서가 개발되기를 기대한다.

현재 연구 결과를 통해 얻을 수 있는 일반적인 시사점은 다음과 같다. 먼저, 기존 과학교과서 텍스트 특성을 분석한 연구는 어절 및 문장 길이와 어휘 빈도 등과 같은 피상적인 수준에서의 특성만 분석하였다면, 현재 연구는 표층 구조, 어휘 정보, 통사적 복잡성 측정치에 더해 명사 반복, 접속부사, 대명사 측정치와 같은 여러 정합성 측정치까지 포함한 보다 다층적이고 심층적인 수준에서 학년 간에 변화하는 텍스트 특성을 보다 포괄적으로 비교 분석하였다. 이를 통해 예상 독자층의 특성에 따라 단계적으로 텍스트 특성이 조절되는지를 다각도에서 분석할 수 있었다. 또한, 기존 과학교과서 텍스트 연구는 한 학년에 국한해서 자료를 분석하거나, 일부 단원에 대한 자료만을 분석 대상으로 한정하였다. 이 때문에 기존 선행 연구에 기반해, 독자 특성을 고려한 텍스트 조절이 단계적으로 이루어지는지를 점검하기에는 한계가 있었다. 현재 연구에서는 중학교 전 학년 및 모든 교과 내용에 대한 텍스트를 추출하여 종합적으로 각 학년 간의 텍스트 특성 차이를 비교분석하였다는 점에서 의의점을 찾을 수 있겠다.

본 연구가 제시하는 이러한 시사점에도 불구하고 본 연구는 추후 연구에서 보완되어야 하는 한계점을 지니고 있다. 이러한 한계점에 기반하여 후속 연구를 제안하고자 한다. 먼저, 본 연구는 중학교 과학교과서 텍스트만을 분석하였다. 텍스트 특성이 상대적으로 크게 영향을 미치는 초중 학습자의 특성을 고려하면, 초등학교 교과서 연구의 필요성이 대두된다. 그동안 초등학교 교과서를 대상으로 한 연구는 어느 정도 수행이 되었지만 대부분 한정된 측정치에 기반해 수행되었기에, 후속 연구에서는 정합성 관련 측정치를 포함한 보다 광범위한 측정치에 기반해 초등학교 과학교과서의 언어적 특성이 보다 종합적으로 비교 분석될 것을 제안한다. 이에 더해, 고등학교 과학교과서에 까지 확장하여 연구가 수행된다면, 과학교과서 텍스트 특성 분석 연구의 토대를 확대하는 데 기여할 것이라 기대된다. 현재 연구에서는 각 학년에 대한 절대적 기준점이 부재한 상황에서 학년 간의 상대적인 언어적 특성을 비교하였는데, 후속 연구에서 각 학년에 기대되는 절대적인 학습자 특성 실태 및 분석 연구가 수행된다면 학습자의 특성에 맞는 언어적 특성을 보다 타당하게 조절될 수 있을 것이라 기대된다.

국문요약

교과서를 통한 학습의 효율성을 최대화하기 위해서는 교과서에 수록된 텍스트 특성이 예상된 학습자의 특성(i.e., 언어적 및 인지적 능력, 배경지식 수준)에 따라 체계적으로 조절되어야 한다. 이에 따라 현재 연구에서는 과학교과서 개발에 이러한 체계적인 원칙이 반영되어 있는지를 알아보기 위하여 중학교 1, 2, 3학년 과학교과서의 학년 간 언어적 특성을 비교 분석하였다. 구체적으로 한국어 분석 프로그램인 Auto-Kohesion 시스템을 활용하여 기존 텍스트 분석 연구에 많이 활용되었던 텍스트 표층 구조 측정치, 어휘 관련 측정치, 통사적 복잡성 측정치와 같은 피상적 측정치에 더하여 여러 정합성 관련 측정치(e.g., 명사 반복, 접속사, 대명사)를 분석하였다. 주요 분석 결과, 대체로 어절 및 문장 길이, 어휘 빈도와 같은 피상적으로 두드러지는 특성에 대해서는 학년이 증가함에 따라 텍스트 복잡도가 상승하는 방향으로 단계적으로 조절이 이루어졌지만, 그 외의 많은 언어적 특

질에 대해서는 체계적으로 조절되지 않은 것으로 나타났다. 특히 여러 정합성 측정치들이 교과서 개발 과정에서 충분히 고려되지 않은 것으로 시사되었다. 이러한 결과는 저학년 학습자들이 교과서를 사용할 때 발달 단계에 맞지 않는 어려운 텍스트를 접할 가능성이 있어서 학습 의욕 및 효율성 저하 현상이 발생할 수 있다는 것을 제시한다. 아울러 고학년 교과서가 고등 교육을 대비하여 더욱 복잡한 텍스트를 처리할 수 있는 능력을 개발시키기 위한 용도로 적절하지 않을 수 있음을 시사한다. 본 연구는, 추후 교과서 개발 과정에서, 예상된 독자 특성의 변화에 따라 정합성 측정치를 포함한 여러 언어적 특성이 단계적으로 조절되어야 함을 제안한다.

주제어 : 과학교과서, 텍스트 분석, 한국어 텍스트 분석, 정합성

References

- Beck, I. L., McKeown, M. G., Sinatra, G. M., & Loxterman, J. A. (1991). Revising social studies text from a text-processing perspective: Evidence of improved comprehensibility. *Reading Research Quarterly*, 26, 251-276
- Berendes, K., Vajjala, S., Meurers, D., Bryant, D., Wagner, W., Chinkina, M., & Trautwein, U. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track?. *Journal of Educational Psychology*, 110(4), 518.
- Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders*, 25(1), 65-83.
- Boscolo, P., & Mason, L. (2003). Topic knowledge, text coherence, and interest: How they interact in learning from instructional texts. *The Journal of Experimental Education*, 71(2), 126-148.
- Caron, J., Micko, H., & Thuring, M. (1988). Conjunctions and the recall of composite sentence. *Journal of Memory and Language*, 27, 309-323.
- Chall, J. S., Conard, S. S. (1991). *Should textbooks challenge students? The case for easier or harder textbooks*. New York, NY: Teachers College Press.
- Chang, H., & Lim, M. (2016). An analysis on suitability of words and sentences in mathematics textbooks for elementary first grade. *The Journal of Educational Research in Mathematics*, 26(2), 247-267.
- Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, 18(3), 257-273.
- Cirilo, R. K. (1981). Referential coherence and text structure in story comprehension. *Journal of Verbal Learning and Verbal Behavior*, 20(3), 358-367.
- Duffy, D. F., Graesser, A. C., Louwse, M. M., & McNamara, D. S. (2006). Assigning grade levels to textbooks: is it just readability?. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 28, No. 28).
- Gasparinatou, A., & Grigoriadou, M. (2013). Exploring the effect of background knowledge and text cohesion on learning from texts in computer science. *Educational Psychology*, 33(6), 645-670.
- Gilbert, R., Martínez, G., & Vidal-Abarca, E. (2005). Some good texts are always better: Text revision to foster inferences of readers with high and low prior background knowledge. *Learning and Instruction*, 15(1), 45-68.
- Graesser, A. C., & Clark, L. F. (1985). The generation of knowledge-based inferences during narrative comprehension. *Advances in Psychology*, 29, 53-94. doi:10.1016/S0166-4115(08)62732-6
- Graesser, A. C., Jeon, M., Cai, Z., & McNamara, D. S. (2008). Automatic analyses of language, discourse, and situation models. In J. Auracher & W. van Peer (Eds.), *New beginnings in literary studies* (pp. 72-88). Cambridge Scholars Publishing.
- Graesser, A. C., Jeon, M., Yan, Y., & Cai, Z. (2007). Discourse cohesion in text and tutorial dialogue. *Information Design Journal*, 15(3), 199-213.
- Graesser, A. C., McNamara, D. S., Louwse, M. M., & Cai, Z. (2004). Coh-Matrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*, 36, 193-202.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Hall, S. S., Maltby, J., Filik, R., & Paterson, K. B. (2016). Key skills for science learning: the importance of text cohesion and reading ability. *Educational Psychology*, 36(2), 191-215.
- Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 13(5), 512-521.
- Kang, S., & Koh, H. (2014). An analysis on the readability of the texts in elementary school science textbooks in terms of word and sentence units. *Journal of Korean Elementary Science Education*, 33(3), 549-557.
- Kang, S., & Park, S. (2009). Comparison of the readabilities of science and social studies textbooks. *Journal of the Society for the International Gifted in Science*, 3(1), 55-65.
- Kang, Y., & Back, S. (2020). A comparative analysis of the word depth appearing in representations used in the definitions of mathematical terms and word problem in elementary school mathematics textbook. *Journal of Elementary Mathematics Education in Korea*, 24(2), 231-257.
- Kendeou, P., & Van Den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & cognition*, 35(7), 1567-1577.
- Kim, B., Ryu, J., & Kang, H. (2005). On optimality text of how well Korean readers fit children in the elementary: with focus on graders from third to sixth. *Journal of Reading Research*, 14, 151-172.
- Kim, J., Park, S., Cho, Y., & Song, S. (2014). Analysis of text difficulty in social studies textbook. *Theory and Research in Citizenship Education*, 46(1), 27-57.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49, 294-303. doi:10.1037/0003-066X.49.4.294
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Koh, H., Song, J., & Kang, S. (2010). A study on the readability of elementary school science textbooks. *Journal of Korean Elementary Science Education*, 29(2), 134-143.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Boston: Addison-Wesley Longman Ltd.
- Lee, J., Maeng, S., Kim, H., Kim, H., & Kim, C. (2007). The systemic functional linguistics analysis of texts in elementary science textbooks by curriculum revision. *Journal of the Korean Association for Science Education*, 27(3), 242-252.
- Lee, S. (2011). A comparative study on readabilities of elementary school textbooks. *Korean Language Education Research*, 41, 169-193.
- McKeown, M. G., Beck, I. L., Sinatra, G. M., & Loxterman, J. A. (1992). The contribution of prior knowledge and coherent text to comprehension. *Reading Research Quarterly*, 27, 79-93.
- McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 55(1), 51.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247-288.
- McNamara, D. S., Graesser, A., & Louwse, M. M. (2012). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini & E. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences*. Lanham, MD: R&L Education.
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1-43.
- Millis, K. K., & Just, M. A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language*, 33(1), 128-147.
- O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional "high-stakes" measures of high school students' science achievement. *American Educational Research Journal*, 44(1), 161-196.
- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, 19(3), 228-242.
- Ozuru, Y., Dempsey, K., Sayroo, J., & McNamara, D. S. (2005). Effects of text cohesion on comprehension of biology texts. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 1696-1701).

- Pearson, P. D. (2013). Research foundations of the Common Core State Standards in English language arts. In S. Neuman & L. Gambrell (Eds.), *Quality reading instruction in the age of Common Core State Standards* (pp. 237-262). Newark, DE: International Reading Association. <http://dx.doi.org/10.1598/0496.17>
- Pressley, M., Wood, E., Woloshyn, V., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychology, 27*, 91-109.
- Ryu, J., & Jeon, M. (2020). An analysis of the continuity of middle school Korean textbooks with Auto-Kohesion. *Language and Linguistics, 89*, 31-60.
- Ryu, J., Lee, H., & Jeon, M. (2020). A comparative analysis of the genres in Korean textbooks using a Korean analysis tool. *Korean Journal of Applied Linguistics, 36*(4), 163-185.
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149-183.
- Suh, H., & Ryu, S. (2014). A study on types and complexity of textbooks. *Korean Language Education Research, 49*(1), 445-470.
- To, V. (2018). Linguistic complexity analysis: A case study of commonly-used textbooks in Vietnam. *Sage Open, 8*(3), 2158244018787586.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.
- van Silfhout, G., Evers-Vermeul, J., & Sanders, T. J. (2014). Establishing coherence in schoolbook texts: How connectives and layout affect students' text comprehension. *Dutch Journal of Applied Linguistics, 3*(1), 1-29.
- Veliutino, E. R. (2003). Individual differences as sources of variability in reading comprehension in elementary school children. In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 51-81). New York: Guilford Press.
- Voss, J. F., & Silfies, L. N. (1996). Learning from history text: The interaction of knowledge and comprehension skill with text structure. *Cognition and Instruction, 14*(1), 45-68.
- Vygotsky L (1976) Play and its role in the mental development of the child. In J. Bruner, A. Jolly & K. Syla (Eds.) *Play, its role in development and evolution* (pp. 537-554). Harmondsworth, Middlesex: Penguin.
- Yun, E. (2019). Analysis the sentences of definition in physics units of science textbooks. *EMLC, 82*, 269-291.

저자정보

류지수(건국대학교 강사)

전문기(건국대학교 교수)