읽기(독서) 교육 체계화를 위한 텍스트 복잡도(Degree of Text Complexity) 상세화 연구 (2)

서 혁 이화여대 국어교육과 교수, 제1저자, 교신저자
이소라 이화여대 국어교육학과 박사과정
류수경 이화여대 국어교육학과 박사과정
오은하 이화여대 국어교육학과 박사과정
윤희성 이화여대 국어교육학과 박사과정
변경가 이화여대 국어교육학과 박사과정
편지윤 이화여대 국어교육학과 석사과정

* 이 논문은 2012년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이다(NRF-2011-327-B00627). 이 논문은 제53회 국어교육학회 학술발표대회(2013. 4. 20)에서 발표한 것을 수정 · 보완한 것으로 2차년도 연구 성과 보고에 해당한다.

I. 머리말

읽기(독서) 교육의 위계화 즉, 학습자 수준에 따른 체계적인 읽기 교육의 실현을 위해서는 텍스트 복잡도(Degree of Text Complexity)[1]를 판단하는 기준을 마련하고 이를 정밀화하는 연구가 필수적이다. 이에 본 연구에서는 텍스트 복잡도 판별을 위한 요인을 체계화 · 명료화 · 객관화하고 이를 바탕으로 한 이독성 공식을 마련함으로써, 학습자의 수준에 따른 체계적인 읽기 교육 시스템 구축에 기여하는 것을 궁극적인 목적으로 삼았다.

텍스트 복잡도와 관련한 연구는 학습자의 발달 단계 혹은 수준에 따라

1 본고에서의 '텍스트 복잡도(Degree of Text Complexity)'란 '텍스트(글)의 복잡성 정도'를 가리킨다. 미국의 중핵 교육과정인 CCSS에서도 교과서 텍스트 수준의 적합성을 고려하면서 텍스트 복잡성(Text complexity)이라는 개념을 논의한 바 있다. 본래 텍스트 복잡도는 텍스트의 내적, 외적 요소를 모두 고려하는 개념이지만 본고에서는 텍스트의 내적 복잡도를 고려한 공식 개발에 초점이 있다. 그간 텍스트 내적인 요소를 다루어 텍스트의 난이도를 설명할 때 '이독성(readability)'이라는 용어를 사용해 왔는데, 이독성은 본고에서 사용하는 텍스트 복잡도의 일부에 속하는 개념이라고 할 수 있다. 텍스트 복잡도 개념에 관한 자세한 내용은 서혁(2011ㄱ)을 참고할 수 있다.

텍스트를 선정 및 제공한다는 점에서, 읽기 교육뿐만 아니라 교육 전반의 수준별 교수 · 학습을 가능하게 하는 중요한 요소로 작용한다. 이러한 텍스트 복잡도 연구가 유의미하게 진행되기 위해서는 텍스트 복잡도 요인을 명료화 · 체계화하는 것이 필수적으로 이루어져야 한다.

미국 공통 중핵 교육과정(The Common Core State Standards for English Language : CCSS)에서는 텍스트 복잡도에 관여하는 요인을 매우 다양하게 제시하고 있는데, 이는 크게 양적 요인, 질적 요인, 독자-과제 요인으로 범주화될 수 있다. 우선 양적 요인에는 단어의 길이나 빈도, 문장 길이, 텍스트의 응집성(cohesion) 등이 포함되며, 질적 요인에는 의미, 글을 읽는 목적, 글의 구조, 텍스트의 언어적 관습 및 글의 명확성, 독자에게 요구되는 지식 수준 등이 포함된다. 독자-과제 요인에는 특정한 독자의 동기, 지식, 경험과 특정한 과제 즉, 목적, 부과된 과제의 복잡도, 제기된 질문 등이 텍스트 복잡도에 영향을 미치는 구체적인 변수로 포함된다(서혁, 2011 ㄱ; 이성영, 2011; 이순영, 2011; 최숙기, 2011 등).

이러한 텍스트 복잡도 요인과 관련한 연구 성과를 바탕으로 텍스트 복잡도 연구에 있어 최근까지 비교적 활발하게 연구되어 온 부분은 이독성(Readability) 연구인데, 이 중 대부분의 연구는 텍스트 복잡도 관련 요인 중 양적 요인을 수량화 · 공식화하는 차원에 집중되어 있다.[2] 기존의 연구들 중에서 지금까지 가장 널리 알려져 많이 활용되고 있는 이독성 공식은 플레시(Flesch) 공식과 데일-챌(Dale-Chall) 공식, 플레시-킨케이드(Flesch-Kincaid) 공식 등을 들 수 있다(서혁, 2011 ㄱ: 434).

심재홍(1991), 윤창욱(2006) 등의 연구는 기존 외국의 이독성 공식을 바탕으로 단어와 문장 수준에서 교과서 텍스트를 이용하여 국어에 맞는 이독

2 현재까지의 국외 이독성 연구는 거닝(Gunning, 1949), 플레시(Flesch, 1948), 데일과 챌(Dale & Chall, 1948 · 1995), 프라이(Fry, 1968), 킨치(Kintsch, 1977), 플레시와 킨케이드(Flesch & Kincaid, 1975), 맥로린(McLaughlin, 1969) 등이 있으며, 국내 이독성 연구는 윤영선(1974), 이선희(1984), 심재홍(1991), 최인숙(2005), 윤창욱(2006) 등이 있다.

성 공식을 개발하고자 한 노작(勞作)의 소산임에 분명하나, 어휘 범주나 분석 기준 및 정확성 측면에서 일정 부분 한계를 지니고 있다.

기존 이독성 연구에서는 다양한 텍스트 복잡도 요인 중에서 양적 요인만을 고려했고, 사실 여러 양적 요인 중에서도 어휘의 난이도 및 문장의 길이 정도에 머물고 있다[3]는 점에서 이독성 공식 및 이를 통해 도출되는 이독성 지수는 불가피하게 한계를 지닌다. 그러나 이러한 불완전함에도 불구하고 이독성 지수는 문식성 진단, 교재 개발, 독서 교육 등의 측면에서 그 현실적인 필요성으로 인하여 여전히 상당한 관심 속에 연구가 진행되고 있다.

이에 본 연구에서는 텍스트 복잡도를 보다 구체적으로 규명하기 위해,

---

3 선행연구들을 살펴본 결과, 대체로 텍스트의 내적 복잡도를 판정함에 있어 이독성 기대 요인으로 어휘 요인과 문장 요인을 고려하고 있음을 확인할 수 있었다. 영문 이독성 공식 약 20개에서 살펴본 요소들을 살펴보면(괄호 안은 각 요인을 반영하고 있는 공식의 수), 어휘요인으로 낱말길이(5), 낱말 난이도(8), 접속어(1), 인칭대명사(3), 서로다른단어(3), 전치사(4) 등을 변수로 보고 있고, 문장 요인으로는 문장길이(12), 단문(1), 구문수치(1), 문장심도(1) 등을 변수로 고려하였다. 문장 요인 중 문장 길이 요인이 전체 18개 공식 중 12개 공식에서 채택되어 이독성에 큰 영향을 미치고 있음을 알 수 있다(E. L. Thorndike, 1921; Kiston & Gray, 1923; Lively & Pressey, 1923; Vogel & Washburne, 1928; Gray & Leary, 1935; Lorge, 1939; Flesch, 1943; Dale & Chall, 1948; Lorge, 1948; Dolch, 1948; Flesch R.E, 1948; Flesch H.I, 1948; Spache, 1953; Gunning, 1958; Powers-Sumner-Kearl, 1958; Yngve, V.H,1960; McLaughlin, 1969; Botel & Granowsky, 1972; Fry, 1977; 윤창욱, 2006: 30-31 참고).
국문 이독성 공식 약 8개에서 살펴본 결과에서는 어휘 요인으로 서로다른단어(1), 낱말난이도(6), 지시어(1), 접속어(1), 동음이의어(1), 전문용어(1), 함축어(1), 인칭대명사(3), 한자어(2), 외래어(1) 등을 다루기도 했다. 다음으로 문장 요인에서는 문장길이(7), 단문(3), 대화문장(1)을 변수로 한 경우가 있었다. 국문 이독성의 공식 역시 문장 길이가 변수로 자주 사용되고 영문과 달리 낱말길이 변수는 유효하지 않았다. 이외에 우리말의 특징인 한자어 등을 고려하려는 시도가 있었으나 이후 선행 연구에서는 그 의미를 크게 평가하고 있지 않음을 파악할 수 있었다(윤영선, 1975; 이선희, 1984; 심재홍, 1991; 김기중, 1993; 김수연, 1994; 최재완, 1995; 전정재, 2001; 윤창욱, 2006: 33 참고).
선행 연구에서는 공식 산출과 계산에 편리한 어휘 요인 및 문장 길이를 위주로 이독성을 판단해 왔고, 이독성을 좀 더 정밀하게 판단할 수 있는 그 외의 요인을 다룬 경우가 후속 연구로 이어지지 못하고 있는 점이 아쉬운 부분이라 할 수 있다.

(1) 이독성 측정을 위한 어휘집 목록 마련, (2) 문장 복잡도 측정을 위한 연구의 주요 내용으로 삼았다. 먼저, 텍스트 복잡도 판별을 위한 어휘를 등급별로 어휘 3천, 5천, 1만 6천 단어군으로 설정하여 기초 어휘로 삼았다. 또한 국내 문장 복잡도 연구로 유일한 김의수(2009)의 연구를 살피고, 실제 문장의 난이도를 반영할 수 있는 문장 복잡도의 계산 체계를 마련하여 텍스트 복잡도 공식에 반영하였다. 여기에 전문가의 텍스트 수준 판별 결과를 포함하여 관련 데이터를 SPSS 프로그램(v 21.0)으로 통계 처리하였다.

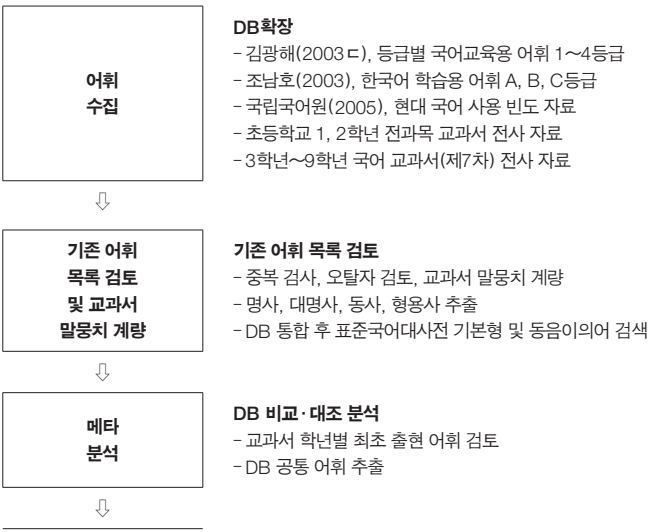
II. 텍스트 복잡도 공식을 위한 어휘 선정

본 연구에서는 기존의 이독성 판별에 쓰인 어휘 목록과 교육용 학습 어휘 선정 관련 연구를 참고하여 3천, 5천, 1만 6천 단어군의 기초 어휘를 선정하였다.[4] 본 연구에서는 교육 장면에서의 실제 어휘 사용 빈도를 추정하여

4 참고로 데일과 챌(Dale & Chall, 1995)에서는 이독성 판별을 위한 쉬운 어휘 목록으로 약 3천 단어를 제시하고 있으며, 윤창욱(2006)에서는 약 5천 단어를 이독성 판별에 사용하고 있다. 본 연구에서는 국어 이독성 판단의 정밀화를 위해서는 기존 이독성 연구에서 사용한 어휘 목록의 확장 및 등급화가 필요하다고 판단하여 3천, 5천, 1만 6천 단어군의 3개의 어휘 등급을 설정하였다. 이와 관련하여 본 연구에서는, 다음의 두어 가지 근거와 관점을 바탕으로, 기존의 Dale-Chall의 3천 단어 기준은 적절하지 않다는 판단을 하였다. (1)Dale-Chall의 이독성 기준은 초등학교 중학년 수준의 기초문식성을 염두에 둔 것으로 판단된다. 그런데 본 연구는 초, 중, 고등학교를 아우르는 국민의 기본 문식성을 염두에 두고 연구를 진행하였다. 이는 국립국어원(2008)의 국민 문해력 조사 연구에서 국민의 평균문식성을 중학교 3학년 수준(국민공통기본교육과정에 의거한 최종 의무교육 기간)에 맞춘 것도 좋은 참조가 된다. (2) 본 연구팀에서도 3천 단어 기준으로 파일럿 테스트를 해 보았으나, 국어 교과서 텍스트의 학년별 복잡도를 명확히 파악하기 어렵다는 결론을 내렸다. 특히 윤창욱(2006)에서도 Dale-Chall(1948) 등에 의거하여 국어 이독성을 적용하여 측정한 결과 3천 단어로는 한계가 있어서 5천 단어까지 상향 조정하여 실시한 것도 좋은 참조가 된다. (3) 영어 이독성 연구는 국어 이독성이나 텍스트 복잡도 연구에 기본적으로

어휘 목록을 마련하였다.[5] 이는 곧 학습자의 연령이나 학령에 따른 어휘 사용 빈도를 고려한 어휘 목록 산출이라고 할 수 있다. 구체적으로, 국어 · 한국어 교육 장면에서 사용되는 어휘들을 대단위로 계량화한 기존의 여러 연구들을 수집하여, 각 연구의 데이터 집합에 제시된 어휘를 교집합한 후 출현 빈도(사용 빈도)가 가장 높은 어휘를 자동적으로 순차화하여 어휘 목록을 마련하는 메타 계량 방식을 사용하였다. 따라서 본 연구의 어휘 목록은 학습자들이 실제로 많이 사용한다고 판단되는 어휘들을 빈도순으로 반영한 것이기 때문에 학습자의 실제 어휘 사용을 고려한 산출이라 볼 수 있다. 연구 절차는 다음과 같다.

표 1. 본 연구의 기초 어휘 목록 작성 절차



도움이 되는 것은 사실이지만, 궁극적으로 접근 방법이나 기준이 달라야 한다는 것이 본 연구 과정에서 내린 결론이다. 단어군에 대한 고려와 구체적인 등급화의 기준은 후술되는 내용을 참고할 수 있다. 이와 관련하여 소중한 의견을 주신 익명의 심사자께 감사드린다.

5 총어휘목록이란 '현대 국어에서 실제로 사용된다고 간주되는 어휘의 목록(김광해, 2003ㄱ: 257)'이라는 관점에서 연구를 진행하였다.

어휘 판정

객관적 기준: 사용 빈도, 중복 횟수
주관적 기준: 학년, 김광해 등급, 조남호 등급, 연구진 판단
서혁(2011ㄷ) 국어 이독성 검사용 쉬운 어휘 목록과의 연속성

기존의 등급별 어휘 목록 중 국어교육용 어휘와 한국어 교육용 어휘를 참조하고, 국립국어원에서 발행한 현대 국어 사용 빈도 말뭉치를 반영하였으며, 교과서 말뭉치를 직접 계량한 것이다. 그리고 이 네 가지 데이터베이스를 전반적으로 비교 대조하는 메타 계량을 실시하고, 단어의 사용 빈도, 교과서 학년, 선행된 등급 판정 등을 종합적으로 고려하여 최종적으로 기초 어휘를 약 3천, 5천, 1만 6천 단어 내외로 마련하였다.

각 등급별 어휘 개수는 다음과 같은 근거를 바탕으로 설정되었다. 앞서 밝혔듯이 가장 공신력 있는 이독성 연구로 꼽히는 Dale & Chall(1995)에서는 이독성 판별을 위한 쉬운 어휘 목록으로 약 3천 단어를 사용하였고, 국내의 윤창욱(2006)에서는 약 5천 단어를 사용하고 있다. 그러나 이와 같이 어휘 목록을 통합적으로 제시할 경우('쉬운 어휘'라는 한 차원으로 제시), 이에 포함되지 않는 단어들에 대한 '난이도' 구분은 해결되지 않을 뿐더러, 오히려 더욱 어려워진다는 한계를 지닌다. 이에 본고에서는 이러한 선행 연구들에서 선정한 어휘 목록 및 그 틀을 발전적으로 수용하여 각각을 어휘 목록의 등급으로 설정하였다. 또한 본고에서 다루고자 하는 어휘 데이터베이스가 선행 연구들에서의 어휘 목록을 메타적으로 분석한 매우 종합적 · 확장적인 것이라는 점, 본고의 연구진들은 기초어휘 목록에는 초등뿐만 아니라 중등학교에서도 두루 사용할 수 있는 어휘까지 확장시킬 필요가 있고, 궁극적으로는 국립국어원(2008)에서 실시한 국민문해력 조사와 같이 전체 국민의 문식성 판단에도 기여할 수 있어야 한다는 점을 고려하여 단어수와 등급을 확장하였다. 이 등급의 어휘 개수는 김광해(2003b)와 본고 연구진의 전문적 판단에 의거하여 약 1만 6천개 내외로 설정하였다.

어휘 수집은 서혁(2011ㄷ),[6] '국어 이독성 검사용 쉬운 어휘 목록'을 마련하는 데에 쓰였던 4가지 데이터베이스인 김광해(2003ㄷ), 조남호(2003), 현대 국어 사용 빈도 자료(김한샘 외, 2005), 교과서 어휘를 기반으로 하여, 추가 확장하는 방법으로 진행되었다. 본고에서의 어휘 데이터베이스 확장 과정은 다음과 같다.

- 김광해(2003ㄷ) 등급별 국어교육용 어휘 1, 2등급 → 1~4등급[7]
- 조남호(2003) 한국어 학습용 어휘 A, B등급 → A, B, C등급
- 1, 2학년 교과서(2007개정) 어휘 → 1, 2학년 교과서(2007개정) 어휘, 3학년~9학년 국어 교과서(7차) 어휘 목록
- 김한샘 외(2005) 현대 국어 사용 빈도 자료 → 동일

김광해(2003)와 조남호(2003)에서는 각각의 어휘집에서 등급을 추가하였고, 교과서 어휘 목록은 교과서 전사 자료를 바탕으로 말뭉치 계량을 실시하였다. 서혁(2011ㄷ)에서는 기초 어휘 목록만 마련하였기 때문에 초등학교 1, 2학년 전 과목에 대한 말뭉치 어휘를 반영하였지만, 초중등학교에서 두루 사용할 수 있는 기초 어휘 선정을 목적으로 등급별 어휘를 1만 단어 수준까지 확장할 필요가 있다고 판단되어, 본 연구에서는 3학년~9학년 국어 교과서 본문 텍스트를 반영하였다.[8]

이에 대한 말뭉치 구축 작업은 다음과 같다. 우선 교과서 전사 자료[9]를

---

6 서혁(2011ㄷ), 고경은(2012) 참조.

7 1등급: 기초 어휘, 2등급: 정규 교육 이전, 3등급: 정규 교육 개시, 4등급: 사춘기 이후(김광해ㄷ, 2003).

8 통계 처리와 관련하여 표본의 확장은 중요한 의의를 지닌다. 그런 점에서 본고 역시 국민기초교육과정인 초등학교 전 교과서의 어휘를 수집하고자 했으나, 여러 가지 어려움으로 초등학교 저학년인 1, 2학년으로 제한하였다. 대신에 7차 국민공통 기본교육과정에 따른 3~9학년 국어 교과서의 전체 어휘를 반영하였다.

9 3~9학년 제7차 국어 교과서 전사 자료는 시각 장애 학생들을 위해 전사된 자료를 활용하

깜짝새 프로그램을 이용하여 어절 단위로 계량하고, 조사와 어미를 삭제하여 단어의 기본형으로 정리하였다. 기본형으로 정리하는 과정에서 표준국어대사전에 등재 여부를 확인하고 동음이의어 번호를 입력하였다. 교과서 어휘를 계량한 결과는 〈표 2〉와 같다. 또한 어휘 수집을 하는 과정에서 기존에 개발된 어휘 목록에 대한 검토를 실시하였다. 검토 결과, 오탈자와 동음이의어 번호 오류[10]를 발견할 수 있었고, 또한 한 단어에 대해 등급이 두 번 부여된 경우[11]가 있었다. 어휘 목록을 검토한 후 4가지 데이터베이스를 통합[12]하였다. 데이터 통합을 통해 단어별로 DB를 정리한 결과를 토대로 표준국어대사전에서 등재 여부를 확인하여 등재되어 있지 않은 어휘를 삭제하였다.[13] 최종 분석에 사용된 어휘량은 〈표 3〉과 같다.

흥미로운 사실은 〈표 2〉의 국어 교과서 어휘 계량에 나타난 바와 같이 학년이 올라감에 따라 국어 교과서에 노출되는 개별 어휘량과 신출 어휘량이 1천 단어 내외 수준에서 지속적으로 증가하고 있다는 점이다.

표 2. 교과서 어휘 계량

어휘 데이터베이스	어휘량 변화	
	검토 전	검토 후
김광해(2003)	33,817	33,663
조남호(2003)	5,858	5,800

였다(출처: E-YAB, www.blind.knise.kr).

10 김광해(2003) 1-4등급에서 '하다 4건', '되다 1건'을 확인하였다. 이는 입력 과정에서의 오탈자이기에 삭제하였다. 조남호(2003)에서는 동음이의어 번호를 일부 수정하였다.

11 '뉴스', '무02'가 김광해(2003) 1등급과 2등급에 중복으로 등재된 것을 확인하였다. 이를 모두 1등급으로 판정하고 2등급에서 삭제하였다.

12 4개 데이터베이스 어휘 목록을 나열해 놓으면 데이터베이스 사이에 중복되는 단어들이 많다. 이에 엑셀의 '데이터 통합' 기능을 통해 한 단어에 대한 서로 다른 출처를 통합하여, 단어별로 데이터베이스 출처를 볼 수 있도록 자동 정리하였다.

13 표준국어대사전 등재 여부를 기준으로 미등재어를 삭제하고 동음이의어 번호를 수정하기도 하였다. 교과서에서 약 1천 단어, 현대 국어 사용 빈도 자료에서 약 7천 단어가 삭제되었는데, 이는 대부분 수사 혹은 등재되지 않은 외국어나 합성어, 파생어였다.

교과서 1-9학년	20,483	19,441
현대 국어 사용 빈도 자료	58,437	51,421
전체	118,595	110,325
데이터 통합	73,244	65,676

표 3. 분석에 사용된 어휘

교과서	개별 어휘량	신출 어휘량
1, 2학년	4,571	4,571
3학년	2,598	1,218
4학년	3,791	1,767
5학년	4,823	2,106
6학년	5,414	2,160
7학년	7,450	2,994
8학년	7,475	2,528
9학년	8,680	3,138
전체	44,802	20,483

이상 각각의 어휘 데이터베이스를 검토한 후 데이터베이스 간에 중복되는 어휘를 검토하였다. 이는 메타 계량의 방법으로 단어 하나하나마다 과거의 연구물들에서 얼마나 많은 지지를 받았었는지 확인할 수 있는데, 비유하자면 이는 곧 단어의 중요도를 '다수결의 원리'에 따라 판정하는 것이다(김광해ㄱ, 2003: 261). 4개의 데이터베이스에서 반복적으로 다루고 있는 단어라면 기초 어휘로 보는 의견이 모아진 것이라 판단하였다.

그러나 4군데 데이터베이스에 모두 출현하는 어휘라고 하여 반드시 기초 어휘라고 할 수는 없다. 가령 4군데 데이터베이스에 모두 출현하는 어휘인 '가치관'의 경우, 사용 빈도는 67회, 김광해(2003ㄷ)에서 4등급, 조남호(2003)에서 C등급, 교과서 최초 출현 학년은 6학년이었다. 데이터베이스가 이미 넓은 등급에 걸쳐 있기 때문에 중복 출현만으로 쉬운 단어라고 할 수 없는 것이다. 즉, 중복 출현은 어휘의 중요도는 설명할 수 있을지 모르나 어

휘의 난이도를 온전히 설명하지는 못한다.

이에 등급별로 3천, 5천, 1만 6천 개의 어휘를 선정하기 위해서는 중복 횟수 외에 또 다른 기준들이 필요하다. 어휘 간의 상대적인 난이도 등수를 세우기 위하여 사용되는 판단의 기준은 일반적으로 말뭉치 계량에 근거한 '빈도'와 '전문가의 판단'에 의존한다. 본 연구에서는 단어 등급화를 위해 객관적 기준으로 '사용 빈도, 교과서 출현 학년, 교집합 횟수'를 상정하고, 기존 전문가들의 판정(김광해ㄷ, 2003 등급; 조남호, 2003 등급)과 더불어 연구진의 종합적 검토와 분석을 통한 메타적인 분석에 따랐다. 마지막으로 본 연구는 서혁(2011ㄷ)에서 개발한 어휘 목록을 바탕으로 하고 있기 때문에 이와 연속성을 유지하기 위하여 서혁(2011ㄷ)에서 개발한 3,510개의 기초 어휘와의 중복 여부도 고려하였다. 이상의 객관적 기준과 주관적 기준에 따라 어휘 데이터베이스를 다양한 방법으로 조합해 본 후, 그 중 등급별 어휘량이 3천, 5천, 1만 6천 단어에 가깝게 나타나는 분포 지점을 찾아내었다. 본래 계획한 3천, 5천, 1만 6천 단어 수를 엄격히 지키기보다는 그와 비슷한 분포가 보일 경우로 유연하게 접근하였다.[14]

- 약 3천 단어 목록(A): 4회 중복∩김광해1,2등급∩초6학년 이하[15]
- 약 5천 단어 목록(B): 서혁(2011ㄷ) 어휘∪4회 중복
- 약 1만 6천 단어 목록(C):
  서혁(2011ㄷ) 어휘∪(김광해 4등급 이하∩사용 빈도 5회 이상)∪(김광해 4등급 이하∩교과서 6학년 이하∩사용빈도 4회 이하)

*A ⊂ B ⊂ C

14 본고의 3천, 5천, 1만6천 단어군 어휘 목록의 정확한 수치는 각각 2,974단어, 5,383단어, 16,040단어이다. 이는 단어 누적 수치에 해당한다.

15 기초 어휘 목록 작성 과정에서 초등학교 6학년 이하 교과서에 등장하는 어휘를 연구진에서 중요하게 고려한 것은 초등학교가 국민 기초 교육기간에 해당하는 시기로 판단했기 때문이다.

A, B, C 사이의 포함관계를 유지하고, 단어 간의 위계를 고려하여 연구진이 단어 수정과 보정 작업을 통해 최종 등급을 판정하였다. 이 A, B, C등급의 등급별 어휘 목록은 기존의 어휘 등급을 참조하고, 현대 언어 사용 빈도와 교과서 어휘 학년을 데이터베이스로 만들어 놓고 3천, 5천, 1만 6천 개로 등급화하였다.[16] 하지만 실제 텍스트 복잡도 판별에 있어서는 이러한 어휘 목록을 융통성 있게 조정하여 적용할 수 있다. 가령 L1 학습자와 L2 학습자에게 제시하는 텍스트는 그 수준이 달라질 수밖에 없다. 그러므로 텍스트 복잡도 판별의 목적에 따라 어휘 데이터베이스를 기반으로 쉬운 어휘 목록을 조정하여 어휘 난이도를 측정할 수 있다.

III. 텍스트 복잡도 공식을 위한 문장 복잡도 체계 개발

1. 문장 복잡도 연구의 현황과 문제점

현재까지 텍스트 복잡도를 판단하는 주요 요인에 대한 논의는 어휘의 난이도나 문장의 길이라는 두 가지 요소에 의존해 온 경향이 컸다. 이는 텍스트 복잡도 연구에서 계속해서 비판받아 왔고, 큰 한계점으로 지적받아 온 부분이다. 이에 어휘의 난이도와 문장의 길이에서 나아가 문장의 복잡도를 추가적으로 고려하는 방안이 제기되었다(서혁, 2011: 440).

문장 복잡도를 수치화하여 텍스트 복잡도에 반영하고자 하는 논의는 다양하게 이루어져 왔다. 문장 복잡도에 관련된 연구를 살펴보면, '통사적 복잡성(Syntactic Complexity)'이란 용어로 다루어지고 있음을 확인할 수 있는데, 이때 통사적 복잡성이란 문장이 생성되거나 변형되는 절차의 복잡성을 의미한다.

16 A, B, C 등급에 속하지 않는 약 5만 여개의 단어들에 대해서는 별도 등급을 부여하지는 않았으나, 잠정적으로 D등급에 해당한다고 볼 수 있다.

문장 복잡도에 관련된 국내 연구는 문장 구조의 철저한 분석에 초점을 두어 이루어지고 있다. 대단위 말뭉치 구축을 위한 방법으로써, 문장 유형의 다양성과 그 구조의 복잡성 측면에서 다루어져 온 해석문법(김의수, 2009)이 그 대표적인 예이다. 문장구조 정보를 문자와 숫자로 표시하는 선형화 모델을 구축하여, 문장의 다양성과 복잡성의 측정기준을 8개 유형, 40개 경우로 구분하여 각각 1점에서 25점까지의 복잡성 점수를 부여하는 체계를 구안했다. 이는 문장의 통사 정보를 드러냄으로써 복잡성의 정도를 분석적으로 제시하고 있다. 가장 단순한 결합이라 할 수 있는 주술 구조를 근간으로, 결합되는 문장 성분의 유형과 개수를 기준으로 복잡성 수치를 산출하는 것이다.

문장 복잡도를 수치화하는 것은 그간 이독성 연구에서 다루어지기 어려웠던 문장 수준의 복잡도를 판단하는 데 큰 도움이 된다. 또한 문장 구조의 다양성과 복잡성은 모문의 차원, 내포문의 차원, 그 둘을 종합한 차원에서 고찰할 수 있다는 다양한 이론적 장점을 지닌다. 김의수(2009, 2010)에서 보여주는 문장 복잡성의 분석은 실로 지난(至難)하고도 '아름다운' 과정이라고 할 수 있다. 다만 분석의 과정이 컴퓨터에 의해 자동화되기 이전까지는 체계적이고도 상세하게 이루어지는 문장 분석의 도식화 과정에 많은 시간이 소요된다. 또한 실제 텍스트 복잡도 판단에 활용하기에는 당장은 몇 가지 아쉬운 제한점을 보여준다. 즉, 먼저 해석문법 체계는 대개 문장 구조를 철저히 분석하는 데 목적이 있었기 때문에 이를 기반으로 제시된 문장 복잡도 점수 체계는 실제 독자들이 느끼는 인지적 부담과 실질적 어려움을 크게 고려하지 못하고 있다. 이는 문장구조의 철저한 분석에 초점이 있는 데서 연유하는 원론적 문제로 이론적·구조적인 측면에서만 문장의 복잡도를 측정한다는 한계를 지닌다. 다시 말해, 구문 분석을 이용하여 구문의 다양성과 복잡성을 측정할 수 있기는 하지만 이것을 활용하여 문장의 복잡도 점수를 산정하는 데는 추가적인 설명이 필요하다. 각 문장 유형의 설정이 이론적으로 제시되어 실제 사례를 찾아보기 어려운 경우도 다수 발견되며, 각 문장 유형의 점수 차이가 얼마만큼의 난이도 증가를 의미하는지도 설명하기 어렵다.

그 밖에도 해석문법을 활용하여 문장 복잡도를 산출하기 위해서는 다음과 같은 문제점 역시 해결되어야 한다. 먼저 분석 기준의 문제이다. 몇몇 문법에 관한 개념이 명확하게 제시되지 않아 분석하는 사람에 따라 해석 결과가 달라질 수 있다. 즉, 평가자의 문법 지식 및 문장 분석의 양상에 따라 복잡도 점수가 달라질 수 있다. 특히 부사절과 부사구의 개념, 이어진 문장과 보조용언의 차이, 보어의 개념 문제 등 논란이 있는 문법 요소들에 관한 분석은 일관된 지침을 가지고 판단을 해야 하는 부분이다. 또한 문장당 최대 점수를 제한하지 않아 그 안에서 내포문의 수가 무한정 늘어나면 한 문장의 복잡도 점수는 무한대로 커질 수 있다. 즉, 문장별로 최대 점수의 격차가 매우 커지게 된다.

본 연구에서는 기존 연구의 장점을 반영하고 한계를 보완하는 방향으로 문장 복잡도를 판단하는 새로운 체계가 필요하다고 보았다. 본 연구에서는 기존의 문장 분석 틀의 수정 및 보완을 거쳐 이를 국어과 텍스트에 확대 적용함으로써 읽기 교육의 체계화 방안에 기여하고자 한다.

2. 문장 복잡도 측정 체계 개발

앞서 언급한 바와 같이 본 연구에서는 텍스트 복잡도 공식 산출에 적합한 문장 복잡도 측정 체계를 새로이 구성하고자 한다. 문장 복잡도에 대한 국내외 논의를 검토하여 적합한 문장 분석 방안 검토 및 대안을 모색하는 것을 목표로 하였다. 이에 국외의 문장 분석에 관한 논의인 보텔과 그래노스키(Botel & Granowsky, 1972)의 연구를 토대로 새로운 문장 복잡도 체계를 고안하였다. 홑문장과 겹문장을 기본으로 하고 수식어 및 절의 안김에 따라 복잡도가 가산되는 원리를 바탕으로 하되, 문장 복잡도 계산에 논란이 있을 수 있는 부분을 찾아 해당 부분에 대해 일관된 분석을 위한 지침을 마련하였다.

보텔과 그래노스키(Botel & Granowsky, 1972)는 어린이들을 대상으로 한 언어수행 연구 결과를 바탕으로 통사적 복잡도 공식을 개발했다. 이는 변형 생성 문법(transformational generative grammar)을 바탕으로 하며, 통

사적 복잡도가 높을수록 어려운 문장이라고 볼 수 있다. 이 공식에서는 문장의 통사적 복잡도를 기본적으로 0~3으로 나누어 살폈다. 그러나 보텔과 그래노스키(1972)의 기준은, 영어와 한국어의 언어 구조상의 차이 때문에 한국어에 바로 적용하는 데는 무리가 따른다.[17] 따라서 본 연구에서는 이를 참고하되, 한국어의 특성에 적합한 문장 복잡도 점수 체계를 여러 차례의 논의와 검토를 거쳐서 다음과 같이 새로 설정하였다.

표 4. 문장 복잡도 분석 체계

문장 단계		문장 형식/ 결합양상	복잡도 (점)
기본 형식		주술	1
		주목술	2
		주보술	
		주목보술/주보목술	3
첨가 조건[18]	①	관형어, 부사어, 독립어 3번 이하	+1
		관형어, 부사어, 독립어 4~6번	+2
		관형어, 부사어, 독립어 7번 이상	+4

17 '주어+동사, 주어+동사+목적어, 주어+be동사+보어' 등 두세 개의 어휘로 이루어진 짧은 문장은 복잡도 점수가 0에 해당한다. '주어+동사+간접 목적어+직접 목적어, 주어+동사+목적어+목적 보어' 등 4개의 어휘로 이루어진 문장은 복잡도 1의 구조에 해당한다. 나아가 수동태, 비교 구문, 분사 구문, 종속절이 있는 문장 구조는 복잡도 점수 2점이 부여된다. 위와 같이 기본적인 문형에 점수를 부과한 뒤 수식어나 수식어구, 조동사, 부정어 등이 첨가되는 경우 첨가조건으로 작용하여 가산점이 부여된다. 이 밖에도 동명사, 부정사, 분사의 용법, 접속사의 종류에 따라 복잡도가 다르게 측정되는 원리이다.

18 '첨가 조건'은 1단계와 2단계의 문형에 첨가될 수 있는 문장 성분을 가리킨다. 그리고 이것이 첨가됨으로써 문장이 복잡해지는 수준에 따라 ①단계와 ②단계로 나뉠 수 있다. 첨가 조건의 ①단계는 보텔과 그래노스키의 공식에서 고려하는 '수식어'와 유사하다고 볼 수 있다. 이는 문장의 복잡도에 ②단계에 비해 문장의 복잡도 수준에 미치는 영향이 미미하다. 그러나 ②단계는 문장의 내포절이 생기게 되는 만큼 문장의 복잡도 수준에 직접적인 영향을 주게 된다. ②단계 내에서도 절의 유형에 따라, 그리고 절과 절의 포함관계에 따라 발생하는 층위로 인해 문장 복잡도 수준은 또 달라질 것이다.

	②	명사절	+3
		관형절	
		부사절	
		인용절	
		서술절	
		절의 수가 6개 이상	+18

위의 표를 바탕으로 문장 복잡도를 산정하는 원리는 다음과 같다. (1) 문장의 기본 형식에 따른 점수를 부여한다. (2) 첨가 조건으로 제시된 ①수식언(독립언)과 ②내포절의 구조를 파악한다. (3) 절이 내포되는 경우 추가 점수를 부여하고, 수식언(독립언)은 절 당 그 개수를 계산하여 추가 점수를 부여한다. 기본적인 점수 산정의 절차에 따른 구체적인 방법과 몇 가지 주요 지침을 제시하면 다음과 같다.

• 기본 형식

- 주술의 구조는 1점, 주목술과 주보술의 구조는 2점, 주목보술의 구조를 지니면 3점이다. 대등하게 이어진 문장의 경우, 기본 형식에 부여된 점수를 산술적으로 합산하여 계산한다.
- '보어'는 서술어 '되다', '아니다' 앞에 격조사 '이/가'를 취하는 성분만을 가리키는 차원(학교문법의 관점)을 넘어서서 필수적 부사어까지 모두 포함하는 포괄적인 개념으로 설정한다. 문장의 분석은 학교문법을 바탕으로 하되, 실제 문장의 복잡도를 고려하여 문장 분석의 기준을 정하였다.

• 첨가조건 ①

- 절 단위에서 관형어, 부사어, 독립어의 유형을 모두 포괄하여 개수를 산정한다. 단, 관형사형 어미 '-ㄴ, -ㄹ', 부사형 어미 '-게' 등 전성어미에 의한 활용형은 '절' 차원으로 다루기 때문에 관형어, 부사어에는 포함하지 않는다. 그 외 일반적으로 관형어, 부사어로 취급하는 것들은 모두 이에 해당한다.

- 명사가 동격으로 나열되어 의미상으로 꾸며주는 경우 해당 명사구를 관형어로 간주하여 첨가 점수를 부여한다. 이는 문장의 복잡도를 판별하기 위해 문법의 형식적 조건보다도 의미 관계를 우선으로 삼은 것이다(ex. '대중 문화의 발생이~'의 경우, '대중', '문화의'를 모두 관형어로 처리함).
- '3번 이하', '4~6번', '7번 이상'의 범주 설정은 연구자들의 사전 조사 분석에 의거한 경험적 판단의 결과이다.

• 첨가조건 ②

- 절이 내포되어 있는 경우 모두 3점을 추가하는 것을 기본 원칙으로 하나, '한 단어'로 제시되는 관형절 또는 부사절(예: 빨간 사과가 맛있다)의 경우 문장을 어렵게 만드는 요인으로 크게 작용하지 않기 때문에 2점만 추가 부여하는 것으로 한다. 한 단어로 이루어진 관형절[19]의 경우는 실제 독자가 느끼는 복잡도(난이도)의 정도가 일반적인 관형절의 경우와 다르기 때문에, 이를 고려하여 '+2'를 부여하는 것이다.
- 첨가조건 ②의 경우 절이 등장할 때마다 절의 성격에 따라 점수를 첨가하고, 이때 절의 기본 형식을 고려하는 것을 주요한 원칙으로 삼는다. 문장의 복잡도는 절 내에 절이 안기는 형태로 그 깊이가 깊어질수록 문장이 어려워진다는 것을 전제로 한다. 따라서 내포절에 다시 내포절이 안긴 경우에는 그 복잡도가 높아짐을 고려하여 절의 유형의 점수를 부여하고 1점을 추가 부여한다.
- 단, 6개 이상의 절이 연속적으로 내포되는 경우 절의 성격에 상관없이 18점을 더하는 것으로 한다. 이는 교과서 문장의 분석상 6개 이상의 절이 안기는 경우는 거의 없지만, 문장의 복잡도가 무한대로 커지는 것을 막기 위한 장치이다.
- 학교문법에서는 부사절과 종속적으로 이어진 문장의 구분은 논란이 되는 부분이다. 본 연구에서 부사절 전성어미를 '-도록, (아)서, -게, -이'로 한정

19 관형절 형성 시 관형절에 해당하는 성분이 관형사형 어미가 결합된 관형어 한 단어로만 존재하는 경우.

하고, 이 외의 경우는 종속적으로 이어진 문장으로 간주한다.

이 기준에 의해 각 문장의 점수를 산정한 후, 한 텍스트의 온전한 문장 분석 점수의 평균을 내면 이것이 곧 한 텍스트(문단)의 복잡도 점수가 된다. 본 연구에서 개발한 문장 복잡도 체계를 기반으로 7차 국어 교과서를 분석한 결과, 학년이 높아질수록 문장 복잡도 점수가 일정 수준 증가하는 경향을 확인할 수 있었다. 학년별 문장 복잡도 변화 추이는 다음과 같다.[20]

표 5. 7차 국어 교과서 텍스트의 학년별 문장 복잡도

학년	1	2	3	4	5	6	7	8	9	10
문장 복잡도	8.37	8.50	9.32	12.65	15.47	15.58	14.43	16.15	19.03	24.08

이처럼 본고에서는 기존의 문장 복잡도 측정 방법에 몇 가지 제한점을 밝히고, 간소화된 문장 복잡도 점수화 방안을 마련하였다. 분석자 간 평가 신뢰도를 높이기 위해서는 실제 활용에 있어 문장 분석의 관점을 통일하고 지침을 명확히 해야 하는 점이 중요한 관건이 된다. 실제 활용에 있어 몇 가지 제약을 극복한다면 기본 구문을 분석하는 데 초점이 있었던 여타 분석 기제보다 문장의 복잡도에 따른 문장 이해의 어려움, 난이도 등을 가장 적합하게 가려내는 체제가 될 수 있을 것으로 기대한다.

본 연구에서 개발한 문장 복잡도 측정 역시 그 과정이 자동화되기 이전에는 직접적으로 문장의 유형과 성분을 분석해 내야 하는 불편함이 있기도 하다. 그러나 그간 이독성 혹은 텍스트 복잡도 공식 산정 시 '효율성'을 추구하여 단어의 난이도와 문장 길이만을 고려해 왔다. 본 연구는 현 상황에 '정

20 분석 대상 텍스트의 선정 기준은 본고의 Ⅳ장에 기술하였다. 학년별 문장 복잡도는 대체로 일정하게 증가하는 경향성을 보여주고 있어 흥미롭다. 다만 7학년의 경우가 6학년보다 낮게 나타나는 예외적 현상을 보여주는데, 이는 향후 더 많은 텍스트 사례들을 적용한 후속 연구가 필요한 것으로 판단된다.

확성'을 보완하려는 목적에서 시도된 것이다. 문장 복잡도를 고려하여 텍스트 난이도를 판단하게 되면, 단순히 문장의 길이로만 텍스트의 어려움을 고려했을 때 판단해 낼 수 없는 문장, 즉 문장이 짧으면서도 어려운 텍스트나 문장이 길면서도 쉽게 느껴지는 텍스트와 관련된 문제들을 문장의 구조 측면에서 상당 수 보완할 수 있을 것으로 기대한다.

IV. 텍스트 복잡도 공식 개발의 과정 및 결과

1. 텍스트 복잡도 공식 개발의 과정

본 절에서는 텍스트의 난이도를 결정하는 데 영향을 미치는 독립변수들을 선정하여 통계 처리함으로써 텍스트의 수준을 측정하는 공식을 도출하는 과정을 제시한다. 텍스트 복잡도 공식 개발의 절차는 〈그림 1〉과 같다.

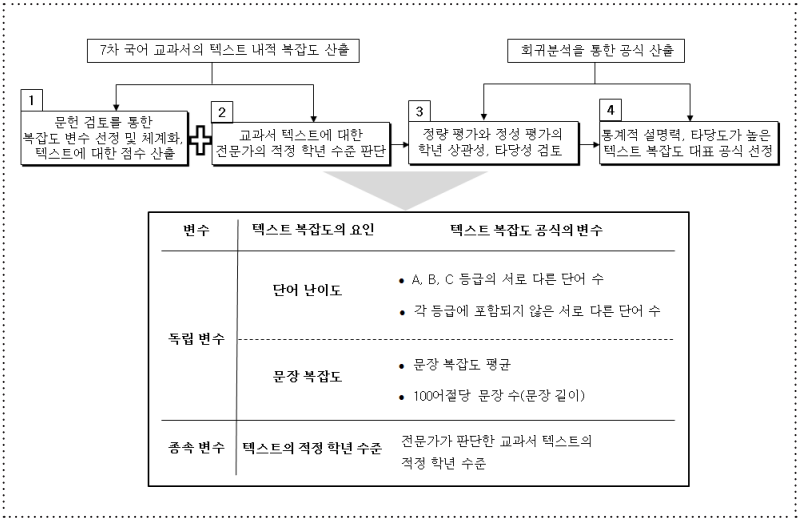


그림 1. 텍스트 복잡도 공식 개발의 과정

먼저 문헌 검토를 통해 복잡도의 변수를 선정하고 이를 체계화하는 단계로 연구를 시작하였다. 기존 문헌을 검토한 결과, 대부분의 관련 선행 연구[21]에서는 텍스트의 이독성이라는 범주에서 텍스트 어휘와 문장의 길이를 주요 변인으로 삼아 공식을 개발하였다. 어휘 측면은 대체로 쉬운 단어의 수나 접속어, 지시어 등을 고려한 것이다. 이러한 맥락에서 윤창욱(2006)에서 제시한 이독성 공식의 설명력은 약 64.7%였다. 본 연구에서는 이러한 기존의 성과들에 대한 메타적 분석을 바탕으로 텍스트 복잡도 공식의 설명력을 높이기 위해 어휘의 난이도를 상세화하고, 문장 복잡도 변수를 새로이 추가하였다.

그 결과 텍스트 복잡도 공식의 요인은 어휘 난이도와 문장 복잡도로 설정하였다. 어휘 난이도 요인의 변수로는 분석 대상 텍스트 내에 나타나는 A, B, C 등급의 서로 다른 단어 수와 각 등급에 포함되지 않은 서로 다른 단어 수를 선정하였다. 문장 복잡도 요인의 변수로는 문장 복잡도 평균과 100어절당 문장 수(문장 길이)[22]를 선정하였다.

변수 선정 이후 이를 적용할 분석 텍스트를 추출하였다. 7차 교육과정 국어 교과서의 1학년부터 10학년까지의 텍스트를 선정하였고, 이를 전문가[23]에게 '적정 학년(수준)'을 판단해 줄 것을 의뢰하였다. 텍스트 선정의 기준은 다음과 같다.

- 국민 공통 기본 교육과정(1-10학년): 읽기 및 국어 교과서
- 텍스트의 종류: 설명적 텍스트(expository text)
- 텍스트의 수: 초등 41개, 중등 48개로 총 89개의 텍스트

---

21 Flesch, 1943; Dale & Chall, 1948; Spache, 1953; 이선희, 1984; 심재홍, 1991; 김수연, 1994; 최재완, 1995; 윤창욱, 2006 등.

22 100어절당 문장 수는 100어절로 이루어진 텍스트 속의 문장 수를 따지는 것이므로, 결국 그 수치가 낮으면 문장의 길이가 길다는 것을 의미한다. 따라서 본 연구에서는 독자의 이해를 돕기 위해 '문장 수(문장 길이)'로 표현하였다.

23 전문가 집단은 국어교육 전공 교수 및 석 · 박사 학위를 소지한 현장 경력 3년 이상의 국어과 교사로, 총 20명(초등 10명, 중등 10명)으로 구성되었다.

• 텍스트의 양: 100어절[24] 기준

위와 같은 텍스트 선정의 과정을 거쳐 연구진은 100어절 기준의 각 텍스트에 대해 ① 어휘의 난이도 등급, ② 문장의 수(문장 길이), ③ 문장 복잡도를 판단하여 데이터를 산출하여 독립변수로 삼았다. 그리고 전문가 집단에게 교과서 텍스트의 수준에 대한 판단을 의뢰한 결과는 종속변수로 반영하였다.

2. 텍스트 복잡도 공식 개발의 결과[25]

수합된 데이터를 바탕으로 SPSS 통계 분석을 통하여 본 연구의 목적에 가장 부합하는 텍스트 복잡도 공식을 산출하였다. 공식 도출을 위한 회귀분석에 앞서, 각 변수들과 전문가들이 판단한 텍스트 적정 학년(수준) 평균과의 상관관계를 살펴보았다. 회귀분석에 앞서 학년(수준) 평균과의 상관관계를 먼저 살피는 이유는, 학년 평균과의 상관이 높은 변수가 회귀분석에서 독립변수로 채택이 되기 때문이다. 상관분석의 결과 초등과 중등에서 학년(수준) 평균에 대한 높은 상관(correlation)을 보이는 변수가 다름을 알 수 있었다.

초등 국어 텍스트에 대한 '학년 평균' 결과와 '단어 난이도, 문장 복잡도'의 상관관계를 살펴보았다.

---

24 선행 연구들에서와 같이 본 연구에서도 전체 텍스트에서 100어절을 추출하고, 이에 대한 등급별 단어 수 및 온전한 문장 수(문장 길이)를 수집하였다.

25 텍스트 복잡도 공식 개발을 위한 통계 처리 과정에서 정확성과 신뢰도를 높이기 위하여 통계 전문가의 자문을 받았다. 이와 관련하여 도움을 준 ㈜와이즈 리서치 김원표 대표께 지면을 빌어 감사드린다.

표 6. 초등 텍스트 '학년 평균'과 독립변수 간의 상관관계

구분		단어 난이도				문장 복잡도	
		A등급 단어 수	A등급 이외 단어 수	B등급 이외 단어 수	C등급 이외 단어 수	문장 복잡도 평균	문장 수 (문장 길이)
초등	학년 평균	-.043	.559**	.555**	-.247	.759**	-.830**

**. 상관계수는 0.01 수준(양쪽)에서 유의

표 7. 초등 텍스트 독립변수 간의 상관관계

구분		단어 난이도	문장 복잡도
		B등급 이외 단어 수	문장 복잡도 평균
초등	A등급 이외 단어 수	.923**	-
	문장 수(문장 길이)	-	-.860**

**. 상관계수는 0.01 수준(양쪽)에서 유의

그 결과 초등 학년 평균과의 상관관계에서 단어 난이도 요인 중에서는 'A등급 이외의 단어 수'가 상관관계가 가장 높았고(0.559, $p < 0.01$),[26] 문장 복잡도 요인에서는 '문장 수(문장 길이)'가 상관이 가장 높았다(-0.830, $p < 0.01$). 'B등급 이외의 단어 수' 역시 'A등급 이외의 단어 수'와 유사하게 상관이 있었으나, 'A등급 이외의 단어 수'와 'B등급 이외의 단어 수' 사이의 상관이 0.923($p < 0.01$)으로 매우 높기 때문에, 회귀분석을 할 경우 'B등급 이외의 단어 수'가 미치는 영향은 'A등급 이외의 단어 수'의 영향력에 의해 상쇄될 것으로 판단하였다. 이는 상관이 높은 두 독립변수를 모두 회귀분석에 투입할 경우 다중공선성 문제가 발생할 수 있기 때문이다. 따라서 두 변수 중 학년 평균에 더 높은 상관관계 및 영향력이 강할 것으로 판단되는 변수를 투입한다. 이는 문장 복잡도의 경우에도 마찬가지이다. 초등 텍스트의 '학년 평균'과 '문장 복잡도'는 0.759로 상관이 높게 나오지만, '문장의 수'가 −0.830

26 이는 기존의 '쉬운 단어 3천 개'의 기준 적용이 국어의 경우 적절하지 않을 수 있음을 시사해 준다.

으로 더 높은 상관을 보이고 있다. 또한 그 두 변수 사이의 상관이 −0.860으로 매우 높기 때문에 초등 텍스트의 학년 평균에 영향을 미치는 문장 복잡도 요인은 '문장 수(문장 길이)'가 더 핵심적인 영향을 미친다고 판단하였다.[27]

표 8. 중등 텍스트 '학년 평균'과 독립변수 간의 상관관계

구분		단어 난이도				문장 복잡도	
		A등급 단어 수	A등급 이외 단어 수	B등급 이외 단어 수	C등급 이외 단어 수	문장 복잡도 평균	문장 수 (문장 길이)
중등	학년 평균	-.445**	.300*	.334*	.380**	.662**	-.568**

**. 상관계수는 0.01 수준(양쪽)에서 유의

표 9. 초중등 텍스트 독립변수 간의 상관관계

구분		단어 난이도	문장 복잡도
		A등급 단어 수	문장 수(문장 길이)
중등	C등급 이외 단어 수	-.253	-
	문장 복잡도 평균	-	-.832**

**. 상관계수는 0.01 수준(양쪽)에서 유의

반면, 중등 텍스트에 대한 '학년 평균'과 독립변수 간의 상관관계에서는 단어 난이도 요인과 문장 복잡도 요인을 측정하기에 가장 상관이 높은 독립변수가 달리 나타났다(표 8, 9 참조).

단어 난이도 요인에 있어서는 'A등급 단어 수'와 'C등급 이외 단어 수'가 각각 -0.445와 0.380으로 유의한 상관을 보였고($p < 0.01$), 문장 복잡도 요인에서는 '문장 복잡도'가 0.662로 '문장의 수' −5.68보다 더 높은 상관을

27 '학년 평균'과 '문장 복잡도'는 정적 상관이고, '학년 평균'과 '문장 수'는 부적 상관으로 상관의 방향이 다르다. 그런데 주지하다시피, 문장 복잡도와 문장수(문장 길이)는 상호 밀접한 관련을 가질 수밖에 없다. 문장의 내포절이 많으면 문장이 길어질 수밖에 없기 때문이다. 회귀 분석에 적절한 독립변수를 확인하기 위해서 상관의 방향은 다르지만, 상관계수의 절댓값을 통해 상관의 크기를 비교하였다.

보였다.[28] 중등 텍스트에서도 역시 '문장 복잡도'와 '문장 수(문장 길이)' 사이의 상관이 높기 때문에 회귀분석에서는 '문장 수(문장 길이)'보다 '문장 복잡도'를 반영하는 것이 적합하다고 판단하였다. 전술한 바와 같이 문장 복잡도와 문장 수(문장 길이)는 상호 밀접한 관련성을 가지기 때문에 한 변인만 반영하는 것이 타당하다고 보기 때문이다. 단어 난이도 요인의 경우 'A등급 단어 수'와 'C등급 이외의 단어 수' 사이의 상관이 -0.253으로 낮고 통계적으로 유의하지도 않기 때문에 두 변수를 모두 회귀분석의 독립변수로 삼기로 하였다.

상관분석 이후 여러 독립변수들을 단계선택법(stepwise)을 적용하여 반복적으로 회귀분석을 실시한 결과에서도 초·중등 텍스트 '학년 평균'에 영향을 미치는 독립변수로 초등의 경우 'A등급 이외의 단어 수'와 '문장 수(문장 길이)'를, 중등의 경우 'A등급 단어 수', 'C등급 이외의 단어 수', '문장 복잡도'를 투입하는 것이 가장 유의하였다.

상관분석과 반복적인 회귀분석 결과를 종합한 결과 텍스트의 '학년 평균'에 영향을 주는 독립변수가 초등과 중등에서 달라짐을 살펴볼 수 있었다. 이는 초등학교 텍스트의 경우 A등급 어휘 목록에 포함되지 않는 어려운 단어가 많을수록, 100어절당 문장 수가 줄어들수록(즉, 문장 길이가 길어질수록) 적정 학년 평균이 올라간다는 것을 알 수 있었다. 반면 중등 텍스트의 적정 학년 평균은 A등급 어휘 목록에 있는 쉬운 단어가 적을수록, C등급 이외의 어려운 단어가 많을수록, 문장의 복잡도가 높아질수록 올라간다는 것을 알 수 있다. 이렇듯 텍스트 '학년 평균'을 설명할 수 있는 독립변수가 초등과 중등에서 달라지므로, 두 집단의 회귀분석을 각각 실시하여 두 개의 회귀선을 찾는 것이 타당하다고 판단하였다.[29]

---

28 상관의 방향은 다르지만 상관 계수의 절댓값을 통해 상관의 크기를 비교하였다.

29 본고는 공식을 개발하는 과정에서 초등과 중등을 아우르는 공식을 개발하는 것보다 초등과 중등의 공식을 따로 개발하는 것이 공식의 정확성을 높이는 데 기여할 것으로 판단하였다. 두 개의 공식을 따로 개발하면서 인상적인 것은 초등과 중등에서 가장 효과적인 변

상관관계 분석을 통해 초등과 중등의 텍스트 '학년 평균'과 상관이 높은 변수들을 중심으로 중다회귀분석을 실시하였고, 이때 변수 투입 방식은 '입력 방법(enter method)'[30]을 선택하였다.

표 10. 초등 텍스트 '학년 평균'에 대한 중다회귀분석(n=41)

독립변수		비표준화 계수		표준화계수	t	유의확률	공선성 통계량	
		B	표준오차	베타			공차	VIF
초등	(상수)	7.438	.733		10.154	.000		
	A등급 이외 단어 수	.032	.017	.173	1.819	.077	.742	1.348
	문장 수(문장 길이)	-.391	.051	-.730	-7.704	.000	.745	1.342
	A등급 이외 단어수 x 문장 수(문장 길이)[31]	.016	.007	.195	2.374	.023	.991	1.009

인이 다르다는 것이었는데, 특히 문장 변인에서의 차이가 눈에 띈다. 초등에서는 문장길이(문장 수), 중등에서는 문장 복잡도가 텍스트 복잡도 판정에 관여하는 바가 더 높다는 것이었다. 이를 구체적인 데이터로 비교해 본 결과, 초등학교 1학년에서 6학년까지 100어절 당 문장 수가 평균적으로 12.75개에서 7.6개로 그 변화하는 추이가 매우 현격함에 비해, 7학년에서 10학년까지의 문장 길이 변화는 7.5개에서 5.3개로 그 변화가 상대적으로 적게 나타났다. 반면 문장 복잡도는 1학년(8.37점)에서 6학년(15.58점)에서의 변화에 비해 7학년(14.43점)에서 10학년(24.08점)에서의 변화 크기가 더 컸다. 이는 중등학년으로 올라갈수록 문장 길이의 변화 폭이 작기 때문에 텍스트 복잡도에 대한 판별력이 낮아지고, 대신 문장의 복잡도가 더 변별력이 높은 변인으로 작용하는 것으로 해석할 수 있다.

30 '입력 방법'이란, 연구자가 선택한 독립변수들이 회귀모형에 동시에 투입되는 것으로서, 독립변수가 어떤 순서에 의해 투입되는 것이 아니라 한꺼번에 투입되어 회귀모형을 결정한다. 따라서 각각의 독립변수는 종속변수를 설명하는 방식에서 다른 독립변수와 공통적으로 설명하는 부분(공통변량)을 제외하고, 각각의 고유한 기여도만을 설명변량으로 갖는다(성태제, 2007: 265).

31 본 연구에서는 독립변수 간의 상관분석과 반복적인 회귀분석을 통해 단어 난이도 단어 난이도와 문장 복잡도 사이에 상관이 있음을 알 수 있었다. 경험적으로 판단해 보건대 학년이 올라갈수록 더 어려운 단어와 복잡한 문장을 구사하게 되는데, 이는 결과적으로 학년 평균에 미치는 '단어 난이도'와 '문장 복잡도'의 영향력이 개별적이라기보다는 계통적으로 관련이 있다고 판단할 수 있다. 이에 두 독립변수의 곱을 상호작용 변수(interaction variable)로 설정하고, 그것의 영향을 상호작용 효과(interaction effect)로 설명하고자 한다.

단어 난이도와 문장 복잡도의 초등 텍스트 '학년 평균'에 대한 기여도와 통계적 유의성을 검정한 결과, 유의수준 0.05에서 '학년 평균'에 유의하게 영향을 미치는 독립변수는 '문장 수(문장 길이)'(t=-7.704, p=0.000)였다. 'A등급 이외 단어 수'(t=1.819, p=0.077)가 종속 변수에 미치는 영향은 유의수준 0.1에서 유의하다. 두 독립변수 간의 상호작용 효과의 경우(t=2.374, p=0.023)0으로 유의수준 0.05에서 유의하다. 초등 텍스트 '학년 평균'을 설명하는 독립변수를 선택하여 그 관계를 나타내는 회귀선을 추정한 결과는 다음과 같다.

ETGL[32]

$= 0.032uW_a - 0.391S + 0.016uW_aS + 7.438$

$* R^2 = 0.752$

ETGL = Elementary Text Grade Level(초등 텍스트 학년 평균)

uW_a = A등급 이외의 서로 다른 단어 수

S = 문장의 수

uW_aS = A등급 이외의 서로 다른 단어 수 × 문장 수

*단어 수와 문장 수는 100어절 기준임.

이 모형은 초등 텍스트에서 100어절당 A등급 이외 단어 수가 1씩 증가함에 따라 학년이 0.32씩 올라가며, 100어절당 문장의 수가 1씩 증가함에 따라 텍스트의 적정 학년이 0.391씩 내려가게 되는 것을 의미한다. 이 회귀모형에 대한 통계적 유의성 검정 결과는 다음과 같다.

---

32 이 식을 풀어서 쓰면 다음과 같다.
초등 텍스트 학년 평균 = 7.438 - 0.391×(문장 수) + 0.032×(A등급 이외의 단어 수) + 0.016×(문장 수×A등급 이외의 단어 수)

표 11. 회귀 모형에 대한 분산분석a

구분		제곱합	자유도	평균 제곱	F	유의확률
초등	회귀 모형	62.588	3	20.863	37.478	.000b
	잔차	20.597	37	.557		
	합계	83.185	40			
R^2(수정된 R^2) = .752(.732)						

a. 종속변수: 학년 평균
b. 예측값: (상수), A등급 이외 단어 수, 문장 수, A등급 이외 단어 수×문장 수

본 연구진이 제시한 초등 텍스트 '학년 평균'을 측정하는 모형에 대한 통계적 유의성 검정 결과, 'A등급 이외 단어 수', '100어절당 문장 수', 'A등급 이외 단어 수×문장 수'가 포함된 모형의 F 통계값은 37.478, 유의확률은 0.000으로 모형에 포함된 독립변수는 유의수준 0.001에서 초등 텍스트 '학년 평균'을 유의하게 설명하고 있으며, 초등 텍스트 '학년 평균'의 총변화량의 75.2%가 모형에 포함된 독립변수에 의해 설명되고 있다.

이러한 결과는, 전술한 A등급 이외의 단어 수의 유의 확률이 0.077로 다소 높다는 제한점을 갖기는 하나, 윤창욱(2006) 등 기존 연구에서 미진했던 동음이의어의 구분 등 어휘 분류의 정밀도를 보강하고, 설명력을 약 10% 포인트 가까이 향상시켰다는 점에서 의의를 갖는다.[33]

다음은 중등 텍스트 '학년 평균'에 대한 중다회귀분석을 실시한 결과이다. 분석에 사용된 독립변수는 상관분석을 통해 선정한 'A등급 단어 수, C등급 이외 단어 수, 문장 복잡도'이다.

33 이와 관련하여 향후 관련 어휘목록과 텍스트 분석의 확장적 적용을 통해 유의 확률을 높여 나가는 연구가 후속될 필요가 있다.

표 12. 중등 텍스트 '전문가 판단'에 대한 중다회귀분석(n=48)

독립변수		비표준화 계수		표준화 계수	t	유의확률	공선성 통계량	
		B	표준오차	베타			공차	VIF
중등	(상수)	9.075	1.015		8.943	.000		
	A등급 단어 수	-.060	.018	-.316	-3.361	.002	.904	1.107
	C등급 이외 단어 수	.145	.056	.238	2.569	.014	.931	1.075
	문장 복잡도	.110	.016	.623	6.787	.000	.950	1.052
	C등급 이외 단어 수×문장 복잡도	.024	.011	.207	2.260	.029	.957	1.045

개별 독립변수의 종속변수에 대한 기여도와 통계적 유의성을 검정한 결과, 유의수준 0.05에서 만족도에 유의하게 영향을 미치는 독립변수는 'A등급 단어 수'(t=-3.361, p=0.002), 'C등급 이외 단어 수'(t=2.569, p=0.014), '문장 복잡도'(t=6.787, p=0.000), '두 독립변수 간의 상호작용 효과'(t=2.260, p=0.029)이며, 유의수준 0.05에서 유의하다. 중등 텍스트 '학년 평균'을 설명하는 독립변수를 선택하여 그 관계를 나타내는 회귀선을 추정한 결과는 다음과 같다.

STGL[34]

$= -0.060W_a + 0.145uW_c + 0.110C + 0.024uW_cC + 9.075$

$^* R^2 = 0.752$

STGL = Secondary Text Grade Level(중등 텍스트 학년 평균)

W_a = A등급 서로 다른 단어 수

uW_c = C등급 이외 서로 다른 단어 수

C = 문장 복잡도

34 이 식을 풀어서 쓰면 다음과 같다.
중등 텍스트 학년
= - 0.060×(A등급 단어 수) + 0.145×(C등급 이외 단어 수) + 0.110×(문장 복잡도) + 0.024×(문장 복잡도×C등급 이외 단어 수) + 9.075
* R^2 = 0.656

uW_cC = C등급 이외의 서로 다른 단어 수 × 문장 복잡도

*단어 수와 문장 복잡도는 100어절 기준임.

이 모형은 중등 텍스트에서 100어절당 A등급에 해당하는 서로 다른 단어 수가 1씩 증가함에 따라 학년이 0.060씩 내려간다는 것을 보여 주고, C등급 이외의 서로 다른 단어 수가 1씩 증가함에 따라 학년이 0.145씩 올라간다는 것을 나타낸다. 또한 문장 복잡도 평균이 1씩 증가함에 따라 학년이 0.110씩 올라가게 된다. 이 회귀모형에 대한 통계적 유의성 검정 결과는 다음과 같다.

표 13. 회귀 모형에 대한 분산분석a

		제곱합	자유도	평균 제곱	F	유의확률
중등	회귀 모형	38.965	4	9.741	20.479	.000b
	잔차	20.454	43	.476		
	합계	59.419	47			
R^2(수정된 R^2) = .656(.624)						

a. 종속변수: 학년 평균
b. 예측값: (상수), A등급 단어 수, C등급 이외 단어 수, 문장 복잡도×C등급 이외 단어 수

세 개의 독립변수로 초등 텍스트 '학년 평균'을 측정하는 모형에 대한 통계적 유의성 검정결과, A등급 단어 수, C등급 이외 단어 수, 문장 복잡도, C등급 이외 단어 수×문장 복잡도(상호작용 효과)가 포함된 모형의 F 통계값은 20.479, 유의확률은 0.000으로 모형에 포함된 독립변수는 유의수준 0.001에서 초등 텍스트 '학년 평균'을 유의하게 설명하고 있었다. 본 모형에 포함된 독립변수들은 중등 텍스트 '학년 평균'의 총변화량의 65.6%를 설명해 준다.

이러한 결과는, 비록 초등과 중등의 텍스트 복잡도 공식을 별도로 제시해야 한다는 번거로움이 있기는 하나, 어휘 분류의 정밀성과 함께 문장 복잡도 요인의 추가 적용을 통해 초등과 중등의 텍스트 복잡도 판별에 대한 고려

요인이 다를 수 있음을 시사하고 있다는 점에서 새로운 논의를 제공하는 것이라 할 수 있다.

V. 맺음말

본 연구는 기존 이독성 공식을 메타적으로 검토하여 보완하고 한국어 텍스트 복잡도 공식을 정밀화함으로써 더욱 체계적이고 객관적인 읽기(독서) 교육을 지향하기 위한 것이다.

본 연구는 보다 타당한 텍스트 복잡도 공식을 개발하기 위하여 텍스트 복잡도 판별을 위한 어휘 목록을 마련하였고, 텍스트 복잡도 판별 요인으로 문장 복잡도 요인을 추가하였다. 이를 바탕으로 회귀분석을 활용하여 초등과 중등 교과서의 텍스트 학년 수준 판별에 적합한 새로운 텍스트 복잡도 공식을 각각 제시하였다. 아울러 초등 수준의 텍스트는 어휘와 문장의 길이 요인이, 중등 수준에서는 어휘와 문장 복잡도 요인이 더 중요하게 작용함을 밝혔다.

본고의 텍스트 복잡도 공식이 상용화되기 위해서는 어휘 목록의 확장과 좀 더 광범위한 텍스트 분석을 바탕으로 한 연구가 축적되어야 한다. 또한 자연언어처리 연구를 기반으로 어휘 난이도 판단과 텍스트 복잡도 계산을 자동화하는 프로그램이 구축되어야 할 것이다.

본 연구에서 마련한 텍스트 복잡도 공식은 초 · 중등학교 교과서의 설명적 텍스트를 중심으로 텍스트 내적 요인만을 중요 변수로 설정하였다는 점에서 일정 부분 한계를 지닌다. 향후 총체적인 텍스트 복잡도 공식 마련을 위해서는 텍스트 복잡도 외적 요인인 배경지식과 과제 요인 등에 대한 후속 연구가 요청된다.

* 본 논문은 2013. 6. 30. 투고되었으며, 2013. 7. 8. 심사가 시작되어 2013. 7. 31. 심사가 종료되었음.

참고문헌

고경은(2012),「이독성 공식의 보완을 위한 학습자 배경지식 위계화 방안 연구」, 이화여자대학교 석사학위 논문.

김광해(2003ㄱ),「국어교육용 어휘와 한국어교육용 어휘」,『국어교육』, 111, 한국어교육학회, pp. 255-291.

_____(2003ㄴ),「메타 계량 방법에 의한 교육용 어휘 선정과 평정」,『계량언어학』, 2, 박이정, pp. 157-357.

_____(2003ㄷ),『등급별 국어교육용 어휘』, 박이정.

김수연(1994),「읽기수준검사 개발을 위한 기초 연구」, 숙명여자대학교 석사학위논문.

김의수 · 이로사(2009),「교육과정에 따른 문장의 다양성과 복잡성 추이 — 중학교 1-1 국어 교과서를 대상으로 —」,『한국언어문학』 제69집, 한국언어문학회, pp. 83-115.

_____(2010),「문장 구조의 다양성과 복잡성」,『시학과 언어학』, 19, 시학과언어학회, pp. 67-97.

김의수 · 정은주(2009),「TOPIK 읽기 영역 지문의 난이도와 균질성에 관한 통사론적 접근」,『한국언어문학』, 71, 한국언어문학회, pp. 189-213.

김의수 · 정한네(2009),「한국어 교재를 구성하는 텍스트의 통사론적 난이도와 균질성 연구」,『어문론총』, 51, 한국문학언어학회, pp. 1-33.

김진경(2012),「통사적 복잡성에 따른 경도인지장애 환자와 알츠하이머성 치매환자의 문장이해능력 비교」, 이화여자대학교 대학원 석사학위논문.

김한샘 외(2005),『현대 국어 사용 빈도 조사 2』, 국립국어원.

김한식(2009),「문장구조의 복잡성과 독이성에 관한 고찰 — 한일 양국의 신문기사문을 중심으로」,『통번역학연구』, 12(2). pp. 145-159.

서 혁(1998),「초등학생의 텍스트성 발달에 관한 사례 연구」,『국어교육학연구』, 8(1), 국어교육학회, pp. 315-349.

_____(2011ㄱ),「읽기(독서) 교육 체계화와 텍스트 복잡도(Text Complexity) 연구(1)」,『국어교육학연구』, 42, 국어교육학회 pp. 433-460.

_____(2011ㄴ),「중등 국어과 교과서 텍스트의 수준과 적합성 검토」,『한국독서학회』, 28, 한국독서학회.

_____(2011ㄷ), '국어 이독성 검사용 쉬운 어휘 목록' (미간행).

성태제(2011),『SPSS/AMOS를 이용한 알기 쉬운 통계 분석』, 학지사.

심재홍(1991),「글의 이독성에 영향을 미치는 요인과 이독성 측정의 모형 연구」, 서울대학교 석사학위 논문.

유혜원(2009),「한국어 구문분석 방법론 연구 — 복문 구조 분석을 중심으로 —」,『민족문화연구』, 50, 고려대학교 민족문학연구원, pp. 153-183.

윤영선(1975),「한국어의 구조적 변인들의 분석과 초등학교 · 중학교 교과서를 중심으로 한

문장 난이도 공식의 개발」, 『연구논문집』, 8, 성신여자대학교 성신인문과학연구소, pp. 209-233.
윤창욱(2006), 「비문학 지문 이독성 공식 개발에 관한 연구」, 한국교원대학교 교육대학원 석사학위논문.
이선희(1984), 「문장 가독성 측정 공식과 이를 통해 본 현대 국어 매스컴 문장의 가독성 측정 조사」, 서강대학교 대학원 석사학위논문.
이성영(2008), 「읽기 발달 단계에 대한 연구」, 『국어교육』, 127, 한국어교육학회, pp. 51-80.
_____(2011), 「초등 교과서의 이독성 비교 연구」, 『국어교육학연구』, 41, 국어교육학회, pp. 169-193.
이순영(2011), 「21세기 국어과 교육과정 개정의 방향 탐색 — 미국의 2010 국가수준교육과정의 특성과 시사점 분석을 중심으로」, 『청람어문교육』, 43, 청람어문교육학회, pp. 7-35.
이정숙(1999), 「통사복합과 이독성과의 관계연구」, 『언어학』, 7(1), 대한언어학회, pp. 361-378.
이지혜(2009), 「Dale & Chall의 이독성 공식을 이용한 한국어 읽기 텍스트 분석 연구」, 경희대학교 석사학위논문.
이혜진(2007), 「문장의 단어 수와 구조를 통한 글 형식 난도 연구」, 경인교육대학교 석사학위논문
이희자(2003), 「국어의 기초 어휘 및 기본 어휘 연구사」, 『새국어생활』, 13(3), 국립국어원.
임성규(1993), 「교재 기술에서 글의 난이도 측정법 연구」, 『한국초등국어교육』, (9), 한국초등국어교육학회, pp. 265-286.
조남호(2002), 『현대 국어 사용 빈도 조사 : 한국어 학습용 어휘 선정을 위한 기초 조사』, 국립국어연구원.
_____(2003), 『한국어 학습용 어휘 선정 결과 보고서』, 국립국어원.
조석주(1982), 「Readability와 Syntactic Complexity」, 『어학연구』, 18(2), 서울대학교 어학연구소, pp. 325-336.
천윤희(2008), 「코퍼스 언어학적 분석을 통한 초 · 중등 영어 교과서의 연계성 연구 : 초등학교 6 학년과 중학교 1 학년 교과서를 대상으로」, 한국교원대학교 석사학위논문.
최숙기(2009), 「중학생의 읽기 효능감 구성 요인 연구」, 『국어교육학연구』, 35, 국어교육학회, pp. 507-544.
_____(2011), 「미국 공통 중핵 교육과정(CCSS)의 읽기 텍스트 위계화 방안에 관한 연구」, 『교육과정평가연구』, 14(2), 한국교육과정평가원, pp. 1-29.
_____(2012), 「텍스트 복잡도 기반의 읽기 교육용 제재의 정합성 평가 모형 개발 연구」, 『국어교육, 139, pp. 451-490.
최인숙(2005), 「독서교육시스템을 위한 텍스트수준 측정 공식구성에 관한 연구」, 『정보관리학회지』, 22(3), pp. 213-232.
최재완(1994), 「신문 경제기사의 讀易性에 관한 연구」, 경희대학교 대학원 박사학위논문.
Botel, M. & Granowsky, A.(1972), A formula for measuring syntactic complexity: A

directional effort, *Elementary English*, 49.
Dale, Edgar & Chall, Jeanne S.(1948), A Formula for Predicting Readability, *Educational Research Bulletin*, vol. 27, pp. 11-28.
Flesch, R.(1943), *Makes of readable style: study in Adult Education*, New York: Teachers College, Columbia University.
Flesch, R.(1948), A New Readability Yardstick, *Journal of Applied Psychology*, 32(2), pp.111-113.
Fry, E.(1968), A Readability Formula That Saves Time, *Journal of reading*, 11(7), pp. 513-578.
Fry, E.(1977) Fry's Readability Graph : Clarifications, Validity and Extension to Level 17, *Journal of Reading*, 21(3). pp. 242-252.
Chall, Jeanne S. & Dale, Edgar(1995), *Readability revisited : the new Dale & Chall readability formul*, Cambridge, Mass. : Brookline Books.
Kintsch, W. Vipond. D.(1977), Reading Comprehension and Readability in educational practice, Paper presented at the conference on memory, university of Uppsala.
Spache, G.(1953), A New Readability Formula for Primary Grade Reading Materials, *Elementary School Journal*, Vol. 55, pp. 410-413.

초록

읽기(독서) 교육 체계화를 위한 텍스트 복잡도 (Degree of Text Complexity) 상세화 연구 (2)

서혁 · 이소라 · 류수경 · 오은하 · 윤희성 · 변경가 · 편지윤

본 연구에서는 읽기(독서) 교육 체계화를 위하여 텍스트 복잡도를 상세화하여, 텍스트 내적 복잡도 공식을 개발하였다. 기존 텍스트 복잡도 공식을 보완하기 위해 (1) 이독성 측정을 위한 등급별 어휘집 마련, (2) 문장의 복잡도 측정 체계 개발을 연구의 주요 내용으로 삼았다.

텍스트 복잡도 공식 개발에 사용된 독립변수는 본고에서 마련한 등급별 어휘 목록 A, B, C등급 및 그 이외의 서로 다른 단어들의 포함 여부와 문장 길이 및 문장 복잡도 측정 체계에 따른 문장 복잡도 평균이다. 또한 종속변수는 전문가가 판단한 텍스트의 적정 학년 평균으로 회귀분석을 통해 텍스트 복잡도 공식을 산출하였다. 이 중 초등에서는 어휘와 문장의 길이가, 중등에서는 어휘와 문장의 복잡도가 더 중요한 변수로 작용하였다. 이를 통하여 초등과 중등 교과서의 텍스트 학년 수준의 적합성을 판별하기에 적절한 텍스트 복잡도 공식을 각각 제시하였다.

본 연구에서 제시하는 초등 및 중등 수준의 텍스트 복잡도 공식은 다음과 같으며, 해당 공식은 초등 텍스트 학년 수준에 대해 75.2%, 중등에 대해 65.6%의 설명력을 지닌다.

초등 텍스트 학년 수준
= 7.438 − 0.391 × (문장 수) + 0.032 × (A등급 이외의 단어 수)
+ 0.016 × (문장 수 × A등급 이외의 단어 수)

* R^2 = 0.752

중등 텍스트 학년 수준
= −0.060×(A등급 단어 수) + 0.145×(C등급 이외 단어 수)
+ 0.110×(문장 복잡도) + 0.024×(문장 복잡도×C등급 이외 단어 수) +
9.075

* R^2 = 0.656

본 연구에서 마련한 텍스트 복잡도 공식은 텍스트 내적 요인만을 중요 변수로 설정하였다는 점에서 일정 부분 한계를 지닌다. 향후 총체적인 텍스트 복잡도 공식 마련을 위해서는 텍스트 복잡도 외적 요인인 배경지식과 과제 요인 등에 대한 후속 연구가 있어야 할 것이다.

핵심어 국어교육, 이독성, 이독성 공식, 텍스트 복잡도, 문장 복잡도, 텍스트 복잡도 공식

ABSTRACT

A Study on the Elaborating the Degree of Text Complexity for the Systematic Reading Education (2)

Suh, Hyuk · Lee, So-ra · Ryu, Su-kyeong · Oh, Eun-ha · Yoon, Hee-sung · Byun, Kyung-ga · Pyeon, Ji-yun

The purpose of this study is to develop Text Complexity Formula more precisely. For this, we focused on complementing the list of words and elaborating the measurement of Sentence Complexity. Independent variables used in this Formula are graded vocabulary list and the figure of sentence complexity calculated by the length and structure of sentence. The dependent variable is the determination on text of the appropriate grade level by experts using regression analysis.

We suggested two Text Complexity Formulars through multiple regression which has the power of explanation of 75.2% about difficulties of texts in elementary level and 65.6% in middle school level.

The Text Complexity Formula, we finally determined, is as follows.

ETGL

$= 0.032uW_a - 0.391S + 0.016uW_aS + 7.438$

$^{*} R^2 = 0.752$

ETGL = Elementary Text Grade Level

uW_a = The number of words except word list A

S = The number of sentence

uW_aS = The number of words except word list A × The number of sentence

STGL

$= -0.060W_a + 0.145uW_c + 0.110C + 0.024uW_cC + 9.075$

$^{*}R^2 = 0.752$

STGL = Secondary Text Grade Level

W_a = The number of words in word list A

uW_c = The number of words except word list C

C = Degree of Sentence Complexity

uW_cC = The number of words except word list C × Degree of Sentence Complexity

This Text Complexity is only focused on the quantitative features of the internal texts. So, it is necessary to conduct follow-up studies for supplementing the formula through including the qualitative factors of the text like Reader variables(reader's background knowledge) and Task variable.

KEYWORDS Korean Education, Readability, Readability Formula, Text complexity, Sentence Complexity, Text complexity Formula