

# 중등학교 학생들의 작문 능력 측정 방안\* \*\*

- 작문 텍스트 분석을 통한 양적 접근법을 중심으로 -

서형국(전북대)

## 〈목 차〉

- |          |             |
|----------|-------------|
| 1. 머리말   | 4. 사례 연구    |
| 2. 기본 사항 | 5. 결론을 대신하여 |
| 3. 분석 방법 | 6. 맺음말      |

## 1. 머리말

이 연구는 중등학교 학생의 평가된 작문 말뭉치를 중심으로 언어 특성을 추출하여 비교함으로써 텍스트를 바탕으로 계량적 평가 가능성이 있는 항목을 추출하여 제안하려는 데 목적이 있다. 특히 본 연구에서는 전라북도 지역의 중등학교 학생들이 실제로 작성한 글을 대상으로, 분석 과정을 실질적으로 제시하고, 이를 현장에서 사용할 수 있는 공개적이고 계량적인 방안으로 개발

\* 본 연구는 2014학년도 제2학기 전북대학교 교육대학원 국어교육과의 전공 강의 <의사소통교육론>의 수강생인 김한나, 강세민, 최은경, 정준철, 박경희, 김보국, 김정효, 김선영, 양은진(자료순)이 수집한 자료를 바탕으로 형태소 분석과 통계 작업을 전연 새로 수행하고, 이를 바탕으로 강의에서 목표한 바를 보고하는 자리를 겸하는 것입니다. 다양한 자료를 모아서 제공하고 강의에서 있었던 다양한 작업을 수행하여 준 수강생의 노고에 감사를 표합니다. 다만, 본 연구에서 제시하는 모든 내용은 발표자의 단독 작업으로 모든 잘못은 발표자에게만 귀속되는 것을 밝힙니다.

\*\* 이 논문은 2014년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.

하여 제시하려고 한다. 나아가 이를 통한 분석의 실체를 실험적으로 제공하려고 한다.

국어를 모국어로 하는 국어 학습자에 대한 측정은 국어 능력에 대한 주관적·객관적 평가를 기반으로 세부적인 탐구가 이루어져야 한다. 국어 학습자의 국어 능력은 교실이나 교육 과정에 의해서만 형성되거나 진행되는 것이 아니기 때문에, 이들에 대한 측정이 교육과정이나 교과서만을 기반으로 하여 이루어지는 데 일정한 한계를 보인다. 따라서 교육과정과 교과서에 대하여 국어 학습자의 국어 능력 측정은 직접적 상관관계를 필연적으로 가진다고 확정하기 어렵다. 그렇다면 국어 학습자의 국어 자료에 대한 주관적·객관적 평가를 통해 세부적인 측정에 대한 기반을 마련하여야 한다.

국어 학습자가 만들어 낸 자료에 대한 주관적 평가는 질적 평가를 통해, 객관적 평가는 양적 평가를 통해 이루어질 수 있다. 질적 평가는 전문가로서 국어 연구자나 국어 교육자가 학습자의 국어 자료를 종합적으로 판단하는 과정을 통해 평가한다. 양적 평가는 일반성을 획득할 수 있는 말뭉치를 대상으로 국어 자료에 대한 변항 중심의 계량이 이루어지는 과정을 통해 평가가 이루어진다.<sup>1)</sup>

국어 학습자가 만들어 낸 자료에 대한 질적 평가는 국어 교육의 제반 단계에서 현재까지 이루어져 왔으나, 평가자의 주관성에 대한 일반화가 어렵다는 단점이 있다. 반면, 국어 학습자가 만들어 낸 자료에 대한 양적 평가는 일반성을 구할 만큼의 대상 자료 확보가 어려울 뿐 아니라, 양적 접근법이 대상 자료에 대한 종합적 평가에 도달할 정도의 측정 수단으로 고려해 볼 만한 자료적 접근이 이루어지지 못하여 왔다.

본 연구에서는 전주 지역 중등학교 학생을 중심으로 이들이 작성한 ‘글’에 대한 양적 접근을 시도하려고 한다. 본 연구에서는 이들 자료를 일정한 양으로 축적하고, 이들에 대하여 양적 접근법을 일관되게 시행하여, 여기서 발견되는 현상을 비교적 사실적으로 제시하려고 한다. 연구의 대상으로서 ‘국어’는 확정적인 대상을 탐구하는 것이라고 할 수 있다(김광해 1994: 61f). 그러나

---

1) 그러나 국어 교육의 직접 대상에 대해서는 양적 평가를 실시하기에 일반성 있는 자료를 구축하기 어려운 문제가 있다. 본 연구는 소규모 자료이기는 하나 세부적인 조사를 통해 비교의 척도를 제안하려는 것이다.

본 연구의 대상이 되는 자료는 상대적이며(Biber & Conrad 2009), 개연적(김광해 1994)이다. 따라서 본 연구의 결과는 상대적, 개연적인 측정을 위한 변향을 계발하기 위한 기초적 연구에 해당한다.

본 연구에서는 양적 접근법의 일반화를 감안하여 다음 몇 가지 점에 유의하여 연구를 진행하기로 한다.

첫째, 공개된 프로그램을 사용한다. 본 연구에서는 비교적 쉽게 구할 수 있는 공개 프로그램을 특정한 수정없이 사용함으로써 재현과 활용이 가능하도록 유도한다.

둘째, 분석 과정에서 엄밀하고 일관된 분석을 실시한다. 다만, 현장에서 학생들의 자료를 평가하는 분들이, 자료에 있을 수 있는 오류를 고치는 외에는, 분석된 자료에 대한 형태 개별적 연구나 분석을 최소화하여 비교 검토하는 과정을 전제하여, 본 조사 결과를 제시한다.

셋째, 전문적 연구 과정이나 추가적인 세부 연구를 위하여 본 연구에서 사용한 방법과 자료에 대하여 재현 가능할 정도의 설명을 부가한다.

## 2. 기본 사항

### 2.1. 자료의 종류

본 연구에서는 2010년 이후에 학생들이 손수 쓴 글을 분석 대상 자료로 삼았다. 학생들의 글은 저학년일수록 길이가 짧고, 고학년일수록 길이가 긴 편이다. 한 편의 길이는 대개 250어절 가량에서부터 800어절 가량까지이다. 각 자료의 길이가 보이는 편차가 언어 능력과 상관관계를 가질 가능성을 감안하여야 할 것이지만, 본 연구에서는 이들을 2500어절 기준으로 모아서 그 평균 수치를 대상으로 판단하기로 한다.

이와 같은 방식의 연구 방법은 몇 가지 문제를 보완하기 위한 것이다.

첫째, 작성자를 통제하는 방식의 글쓰기를 수행하지 않으면서도, 대상 자료를 일정하게 유지할 수 있게 된다. 우리의 연구 자료는 연구자가 자료의 구성에 개입하는 과정을 최대한 억제함으로써 이루어진다. 따라서 자료의 길이나

내용에 대한 통제를 가하지 않고 이루어진 자료에 대한 검토를 통해 작성자의 언어 능력을 검토하고 평가하기 위한 것이다.

둘째, 말뭉치 자료를 대상으로 하는 평균적 말뭉치의 길이와 초·중등학교 학생들이 작성하는 글의 길이에서 나타나는 상위(相違)를 극복하기 위한 방안으로 자료를 집적하는 과정을 거치기로 한다. 대용량 말뭉치를 대상으로 하는 언어 연구에서는 표본 추출을 통하여 연구의 용이성과 신뢰성을 확보하였다.<sup>2)</sup> 이와 같은 방식에서 사용한 표본의 길이는 2,000어절부터 10,000어절 정도에 해당한다. 그러나 본 연구에서 수집한 자료는 최대 1,000어절을 넘지 못하였다. 이를 바탕으로 자료들 간에 보이는 편차를 극복하고 신뢰성을 얻을 수 있는 방안은 이들을 전체로 집적하여 일정 어절이 되도록 하는 방안을 선택하는 것이 합리적이라고 판단하였다.<sup>3)</sup>

본 연구에서 사용한 자료는 다음 <표 1>과 같다.

<표 1> 자료의 목록과 자료량

학 년	구 분	담화 상황				주제	출처	어절 수	형태 수	비고
		성별	줄거리	주장	감정					
5		混	×	○	×	양성평등	글짓기 대회	2,830	5,755	
7	①	여	○	×	○	독서감상	과제	1,613	3,209	교열됨
	②	여	×	×	○	감사인사	사신(私信)	2,585	5,082	편지
8		混	○	×	○	한가위	손글씨 과제	2,862	5,730	
9		混	○	×	×	행사소감	과제	2,098	4,398	
11	①	여	○	×	○	감상문 (‘추억’)	글짓기 대회	2,701	5,492	
	②	남	×	△	○	자유주제	글짓기 대회	2,852	5,759	
12	①	여	×	×	×	자기소개	과제	2,510	5,366	
	②	여	○	×	○	기행문	과제	2,358	4,774	
계								20,051	40,791	

2) 안예리·이주현(2014)에서는 길이가 긴 텍스트(이 경우, 『빈상설』)를 임의의 구간에서 일정량(2500어절)을 추출하여 검토할 수 있는 방안을 제시하였다. 이런 연구방법을 택하게 된 근거가 될 만한 해외의 표본 추출 방식으로는 Helsinki 말뭉치의 경우와 16~18세기 영어 대화 말뭉치인 CED 말뭉치의 경우를 예시하였다.

3) Biber(2012)에서는 포괄적 언어 특성의 경우에, 온전한 텍스트를 전체로서 검토하거나 표본이 되는 텍스트를 모이든 집합을 검토함으로써 레지스터 확인이 가능하다는 점을 지적하였다(특히 192f 참고).

본 연구에서 시도하는 ‘온전한 자료 전체’를 연구 대상으로 하면서, 일정한 분량이 될 때까지 이들 자료를 집적하는 연구 방법은 앞서 언급한 장점, 또는 불가피한 연구 방법에도 불구하고 몇 가지 문제를 보인다.

글의 길이는 그 자체로 언어 능력을 다면적으로 반영한다. 글쓰기를 배우는 수준의 학생일수록, 즉 고급 수준의 글쓰기가 아닐수록, 대개 글쓰기 능력과 글의 길이는 비례하는 것으로 보인다. <표 1>의 글로만 보아도 5·7학년의 글이 극단적으로 짧고, 12학년의 글이 평균적으로 가장 길다. 나아가, 글의 도입 부분이나 맺음말 부분 등 글의 위치에 따라서 레지스터는 일정한 특성을 보일 가능성이 높는데, 글의 길이가 짧은 글을 모은 자료일수록 이러한 위치에 따른 레지스터에 영향을 받는 부분의 비중이 늘어날 수밖에 없다.

이와 같은 문제를 극복하는 방안은, 일정한 수준에 있는 글을 모아서 이들끼리 상호 비교를 통하여 상대적인 평가를 수행하는 것이다. 일반적으로 말뭉치를 활용한 연구에서는 자료의 특성을 규정하여 주는 특성을 일방적으로, 또는 절대적으로 제시하는 것이 무척 어렵다(Biber & Conrad 2009). 그보다 더 자연스럽고 타당한 방법은 대상 자료의 맥락 특성(‘상황 특성’, situational characteristics)을 일정하게 한정하고, 이들 자료에서 특성을 추출하여 상대 비교를 수행하는 방안이다. 본 연구는 이와 같은 연구의 한 방안으로 전주시역 학생의 글쓰기 자료를 바탕으로 언어 특성을 추출하여 보이려는 것이다.

<표 1>에서 보인 자료는 모두 학생이 쓴 글로서 다음 <표 2>를 기준으로 그 자료적 특성을 가늠할 수 있다. <표 2>의 분류 방식으로 씌으로써 국어 자료를 그 생산 맥락을 감안할 수 없을 상황까지를 고려하여 특성별로 분류할 수 있을 것이다.<sup>4)</sup>

---

4) 아래 자료는 줄고(2014)에서 제시한 국어사 자료의 텍스트 유형 확인을 위한 자료 분류 기준에서 약간의 수정을 가하여 제시한 것이다. 문자언어로 제시된 자료를 작성자 여건보다 텍스트 자체에 집중하는 유형 분류 방식이므로 현대국어 공식적인 자료에 적용하는 데 큰 무리가 있지는 않을 것으로 기대한다. 2. 담화 맥락의 경우, ‘공공연함’을 ‘과시적임’으로 바꾸었다. 이때 ‘과시적임’이란 ‘다른 여러 사람에게 보일 것을 전제로 하는 담화’를 특정하기 위한 것이다. ‘공공연’이란 그것이 외부에 공개되었음을 가리키는 결과적 상황에 해당할 뿐, 담화의 발생 맥락을 뜻하지 않기 때문이다. 이런 담화 텍스트의 특성에 따라서 이른바 ‘서사시제’(이지영 2002)가 지목된 바 있는데, 우리는 이 외에도 ‘서사시제’와 유사하게 ‘서사대우’의 가능성을 짚어두기로 한다.

〈표 2〉 국어 자료의 텍스트 유형 확인을 위한 자료(genre) 분류 기준

연번	기준		해당 자료	배제 자료
1	담화 유형	대화인가?	대화, 독백	산문, 운문
2	담화 맥락	과시적인가?	공연, 공간 자료	언간
3	담화 참여자	화자·필자는 명시적인가?	언간, 대화	서술문, 협주
4		청자·독자는 명시적인가?	언간, 대화	서술문, 협주
5	사회 관계 맥락	존대 상황인가?		
5'		하대 상황인가?		
6	작성자 맥락	계획적 담화인가?	간본	편지
6'		교열된 자료인가?	간본	
7	메시지 형식	형식에 제약이 있는가?	운문, (관소리)	
8	메시지 내용	일상적 내용인가?	언간	
8'		상상적 내용인가?	고소설, (傳)	

## 2.2. 자료의 속성

〈표 2〉에 근거하여 〈표 1〉의 자료를 검토하면 다음과 같이 우리의 분석 자료가 보이는 담화 맥락의 특성을 다음과 같이 정리할 수 있다.

첫째, 본 연구의 대상이 되는 자료는 대화가 아니다. 이른바 상관적 장면의 텍스트가 아니라 단독적 장면의 텍스트가 주로 대상이 된다. 따라서 (상대)높임법과 시제의 사용에서 상관적 장면과 다른 특수성이 발견될 것이다.<sup>5)</sup>

둘째, 본 연구의 대상이 되는 자료는 과시적인 자료이다. 즉 작성자 아닌 다른 사람에게 이 글을 보이고 평가받을 것을 전제하고 작성한 자료이다. 이와 관련하여 〈표1〉에서 ‘사신(私信)’이 가지는 특수성은 여타의 ‘과시적 자료’와 다른 속성을 가진다. 이 자료는 다른 자료에 비하여 공개적이지 않은 글쓰기에 해당한다. 또한 공식적이지 않은 글쓰기에도 포함된다.<sup>6)</sup>

5) 이러한 특성으로 인하여 중등학교 학생의 작문 텍스트는 말하기나 듣기 능력보다 읽기 능력이나 독서량과 직접 상관될 가능성이 더 크다. 즉 말하기나 듣기 능력이 언어 능력의 공통 부분을 제외하면 — 우리는 여기에 어휘 능력과 문법 능력이 포함된다고 본다 — 문어(文語)나 글을 대상으로 하는 해득력, 구사력이 특히 주요하게 관련될 것으로 판단된다.

6) 따라서 〈표1〉의 ‘사신’은 본 연구의 자료와 별도로 처리하여야 할지 모른다. 이 원고를 검토하여 준 장은주(교원대) 선생 역시 이와 같은 문제를 지적하였다. 본 연구에서는 7학년의

셋째, 본 연구에 소용된 자료는 대개 작성자가 매우 구체적이지만, 내용에서 그것이 명시적으로 드러나지 않는다. 또한 수신자도 대개 구체적이지만, 명시적이지도 않다. 다만 몇몇 글에서는 작성자가 상당히 구체적으로 명시되기도 한다. 또한 편지글에서는 수신자도 매우 구체적으로 명시되어 있다.

넷째, 이들 자료는 계획적 자료이지만, 대개 즉흥적으로 작성되며, 교열은 대개 이루어지지 않았을 것으로 예상된다. 본 연구의 대상 자료는 대개 ‘과제’를 통해 작성된 글이다. 그러나 실제로 자료 검토 과정에서 오자와 탈자 등이 전반적으로 교열된 수준에 이르렀다고 보기 어려웠다. 따라서 의미 그대로의 ‘교열’을 거친 자료와는 다르게 평가될 필요가 있다고 판단한다.

다섯째, 본 연구의 대상 자료는 형식에 제약을 받지 않는 산문 자료이다. 본 연구의 대상 자료는 대개 ‘주제’가 주어진 글로서 형식적 특수성이 제한되어 있지 않거나, 그 특수성이 변별적으로 구현되었다고 판단하기 어려웠다.

여섯째, 본 연구의 대상 자료는 모두 비전문적 내용을 다루고 있다. 본 연구의 대상 자료는 일부 일상적이지 않은 주제를 다룬 경우조차도 대개 내용 자체를 집중적으로 다루지 못하고 독서 행위나 과제 수행의 외적 과정을 비중 있게 다루으로써 내용의 비전문성을 반영하고 있다고 판단하였다.

마지막으로 본 연구의 대상 자료는 ‘글’이다. <표 2>는 국어사 자료를 대상으로 하였으므로 ‘말’, 즉 ‘말’을 말뭉치화한 전사 자료인지, ‘대본’인지, ‘글’인지 등은 고려 대상이 아니었다. 그러나 이에 대하여 본 연구의 대상 자료는 모두 ‘글’이라는 점을 확인하여 두기로 한다.<sup>7)</sup>

### 3. 분석 방법

본 연구에서는 연구 대상 자료를 다음과 같은 절차를 거쳐 형태 분석과 형태소 분석을 수행하였다. 텍스트 자료는 전자 자료로 구성하고, 문화관광부에

---

과시적 글쓰기 자료가 별도로 존재하는 점을 감안하고, ‘사신’이 갖는 특수성이 있음을 고려하여 계량하여 검토하기로 한다.

7) 본 연구의 과정에서 수집한 자료 가운데 ‘말’을 녹취한 자료가 두 건이 존재하였다. 그러나 본 연구에서는 우선 학생이 쓴 ‘글’에 집중하고, 이들 ‘말’ 자료는 후후의 연구로 미루어 두기로 하였다. 관련 수강생들의 양해를 구한다.

서 2001년에 개발하여 보급한 지능형 형태소 분석기 2.0을 이용하여 형태 분석을 하되, 본 연구의 일관된 기준에 따라 다음과 같이 수정하였다. 또한 이렇게 분석한 자료는 SynKDP 1.5.5로 빈도를 추출하였다.<sup>8)</sup>

본 연구에서 사용한 자료의 표지 단자(tag set)는 다음 <표 3>과 같다.

<표 3> 형태 분석 표지와 약호

대분류	소분류	세분류	대분류	소분류	세분류	
체언	명사 NN	일반명사 NNG 고유명사 NNP 의존명사 NNB	관계언	격조사 JK	주격조사 JKS 보격조사 JKC 관형격조사 JKG 목적격조사 JKO 부사격조사 JKB 호격조사 JKV 인용격조사 JKQ	
	대명사 NP					
	수사 NR					
응언	동사 VV				보조사 JX	
	형용사 VA				접속조사 JC	
	보조용언 VX				의존 형태	어미 E
지정사 VC	긍정지정사 VCP 부정지정사 VCN	접두사 XP	체언접두사 XPN			
수식언	관형사 MM		접미사 XS	명사파생접미사 XSA 동사파생접미사 XSV 형용사파생접미사 XSA (부사파생접미사 XSB)		
	부사 MA	일반부사 MAG 접속부사 MAJ		어기 XR		
독립언	감탄사 IC					

본 연구에서는 대상 자료를 일부가 아닌 전체(complete texts)로서 비교의 대상으로 삼아서, 형태소에 대한 계량적 방법을 사용하여 이들 자료에 대한 비교를 수행하려고 한다.<sup>9)</sup>

8) SynKDP('깜짝새')에서 빈도를 추출하는 경우, 자모로 시작하는 어절은 어절 빈도에서 누락 되는 오류가 발생한다. 본 연구에서는 이 문제에 대응하기 위하여 자모로 시작하는 형태(예컨대, 종결어미 '-ㅂ니다', 명사형 어미 '-ㅁ' 등)는 별도로 계산하여 최종 자료에 합산하였다.

9) 그러나 이와 같은 양적 접근법이 갖는 한계는 국내외를 막론하고 누차 지적이 있어 왔다. 대표적으로 Ullmann의 지적을 인용한 김상태(1982: 92-93)의 요점을 들어 두기로 한다.

이 절에서는 본 연구에서 형태 분석과 형태소 분석을 위해서 세운 기준을 제시하여 두기로 한다.

### 3.1. 형태 분석의 원칙

본 연구에서 시행한 형태 분석의 원칙으로서 앞으로의 연구에서 시도할 만한 원칙은 다음 (1)~(3)과 같다.

- (1) 『표준국어대사전』(이하 ‘<표준>’)을 기준으로 일관된 처리를 하는 것을 원칙으로 한다. 그러나 전체적으로 ‘담화 기능’을 감안하여 『고려대한국어대사전』(이하 ‘<고려>’)과 같은 최근의 대사전류의 처리 결과도 아울러 검토하고, 필요에 따라 <표준>과 다르게 처리하기도 하였다. 예컨대, ‘그렇다면’은 부사로 처리하였는데, 이를 <표준>은 활용형으로 처리했으나, 담화 기능상으로는 ‘대응’의 의미 기능보다 이른바 ‘접속’의 기능을 하는 형태적 단위로 확립된 것으로 보았다. <고려>처럼 ‘부사’로 처리한 경우가 있는 점도 참고하였다.<sup>10)</sup>
- (2) 자료에 쓰인 형태가 같더라도 맥락에 따라 달리 해석하여야 하는 경우, 맥락에 맞는 형태 분석 결과를 제시하였다. 예를 들어, ‘-니다고’의 경우 다음과 같이 그 기능이 서로 다른 사례가 발견된다. (ㄱ)의 경우는 종결어미 ‘-니다’(EF)에 인용격조사 ‘고’(JKQ)가 통합한 사례이고, (ㄴ)의 경우는 연결어미 ‘-니다고’(EC)의 사례이다.  
ㄱ. 마침 고등학교생들이 영자 책자 제작을 한다고 해서 참여했습니다.(12-①)  
ㄴ. 셋째 이모부는 일 때문에 가봐야 한다고 용돈을 주시고 가셨다.(8)
- (3) 분석 도구는 문화관광부에서 2001년 1월에 배포한 <지능형형태소분석기

- 
- ㉠ 문체의 민감한 뉘앙스를 붙잡기에 너무 거칠.
  - ㉡ 데이터에 대한 일종의 허위의 정확성을 줄 수도 있음.
  - ㉢ context의 영향에 대한 대비책이 없음.
  - ㉣ 질과 양에 압도당할 위험이 있음. 즉 다양한 요소가 피상적인 동질성의 바탕에서 함부로 분류당할 위험이 있음.
  - ㉤ 뻔히 아는 결과를 산출하여 낼 때가 있음.

10) 우리는 이와 같은 원칙을 적용함으로써 <표준>을 기계적으로 적용하기보다 실제 학생들이 작문이나 독서 과정에서 문법 단위에 대하여 내린 판단을 실제 텍스트 분석에 적용하는 것이 바람직하다고 제안한다. 이런 원칙은 <표준>을 따르는 방식이기는 하지만 특히 (2)의 원칙에서 더욱 선명하게 드러난다.

2.0>을 사용하고, 대개의 형태 분석은 일반적이면서도 기본적인 원칙에 따랐다. 이에 따라 파생어인 용언은 접사를 분석하는 것을 기본으로 하였으나, 합성용언, 체언복합어는 분석을 하지 않는 것을 원칙으로 하였다. 이는 파생용언의 경우는 특히 어근과 통합하는 사례를 분석할 수 있도록 하기 위함이지만, 일반 명사에 동사나 형용사 파생 접미사가 통합한 사례는 어쩔 수 없이 명사의 분석 결과에서 과분석의 결과가 도출되었다.

### 3.2. 형태소 제시의 원칙

본 연구에서 통계는 형태소별로 제시하는 것을 원칙으로 하였다. 따라서 2.2.1에서 분석한 형태 분석의 결과는 형태소별로 통합함으로써 형태소 분석의 자료로 변환하는 것이 중요하였다. 본 연구에서 적용한 형태소별 분류의 원칙은 다음과 같다.

- (4) 형태 분석의 결과는 형태소별로 대표형을 정하여 제시한다. 이 과정은 분석 자료에 대한 정확한 포집을 위해 선결되어야 하는 과정이다. 이에 따라 아래의 부수 원칙이 수립되었다.
  - ㄱ. 매개모음을 가진 형태는 매개모음을 가진 형태를 대표형으로 삼는다. ‘-나다’와 ‘-은다’는 ‘-은다’로, ‘-으며’와 ‘-며’는 ‘-으며’로 제시하는 따위이다.
  - ㄴ. ‘란’과 ‘이란’처럼 ‘이-’를 매개로 하는 형태의 경우도 긴 형태인 ‘이란’을 대표형으로 삼는다.
  - ㄷ. ‘와, 과’의 경우에는 음성적으로 무거운 ‘과’를 대표형으로 삼는다.
  - ㄹ. 다만, 매개모음을 가진 형태로 간주될 수 있는 경우라도 ‘을, 를, ㄹ’의 경우에 ‘ㄹ’은 ‘를’에 포함하여 제시한다.
- (6) 음성모음형을 대표형으로 삼는다. ‘-어도’와 ‘-아도’ 가운데는 ‘-어도’를 대표형으로 삼는 따위이다.
- (7) ‘ㄷ~ㄹ’의 교체는 별개 형태소로 제시한다. 종결어미 ‘-다’와 ‘-라’의 분포는 서술격조사의 뒤라는 명확한 분포 요건이 존재하므로 한 형태소로 보는

것이 타당하나, 이들을 별개 형태소로 제시함으로써 서술격조사와의 결합 빈도 확인이 가능하다는 장점이 있다.

- (8) 관형사형의 ‘-은’은 선행하는 품사에 따라 동음어적 속성을 띠지만, 본 연구에서는 이를 구분하여 제시할 여력을 가지지 못하였다.
- (9) 형태소별 제시가 무의미한 것은 형태별로 제시한다. 예컨대, 접속조사(JC) ‘과’와 ‘와’라든지, 목적격조사(JKO) ‘을’과 ‘를’은 대개 배타적 분포를 보이는 경우가 많고, 여기에 다른 조사가 첨가되더라도(JC ‘하고’, ‘랑’ 등), 그 분포를 참고하기 쉬우므로 이들은 형태별로 제시한다.
- (10) 선행 형태소가 음운론적으로 매우 직관적으로 이형태 출현 조건으로 성립하는 경우는 통합하여 제시한다. 주격조사(JKS) ‘이’와 ‘가’의 경우, 종결어미(EF) ‘-히니다’와 ‘-습니다’, 종결어미(EF) ‘-는다’와 ‘-니다’는 그 분포를 별도로 따지는 것이 특별한 의미를 부여하지 못하므로, 이들은 모두 밑줄 친 형태를 대표형으로 하여 제시한다.
- (11) 같은 범주를 대표하더라도 음운론적 조건에 따른 이형태가 아닌 경우에는 대표형으로 통합하여 제시하지 않는다. 종결어미(EF) ‘-다’와 ‘-니다’는 범주상으로는 직설법이나 중화된 종결어미의 범주를 대표하지만, 선행하는 용언의 품사가 달라 그 형태를 통합하는 것보다 분리하는 것이 유용성이 있으므로, 이들을 통합하여 제시하지 않는다.
- (12) 준말로 처리될 만한 조사는 본디말이나 중복형을 대표형으로 선정하여 제시한다. 즉, 보조사(JX) ‘는’과 ‘나’은 ‘는’을 대표형으로 삼고, ‘에서’의 준말로 처리된 ‘서’(‘여기서’의 ‘서’)는 ‘에서’에 통합하여 제시한다.

#### 4. 사례 연구

이 장에서는 앞서 제시한 방식에 따라 분석한 자료를 대상으로 크게 어휘 부류와 문법 부류로 나누어 이들이 보이는 양상을 서술하고, 그것이 어떤 특징적 현상으로 귀납될 수 있는지 살펴보기로 한다. 본 연구에서는 실제 자료 분석을 통하여 각 자료의 분석을 통해 시범적인 양적 분석을 실시하고, 이 결과를 확정적 사실 판단이 아니라 앞으로의 양적 분석에서 변항으로 적용할 수 있을 가능성을 제안하기로 한다.<sup>11)</sup>

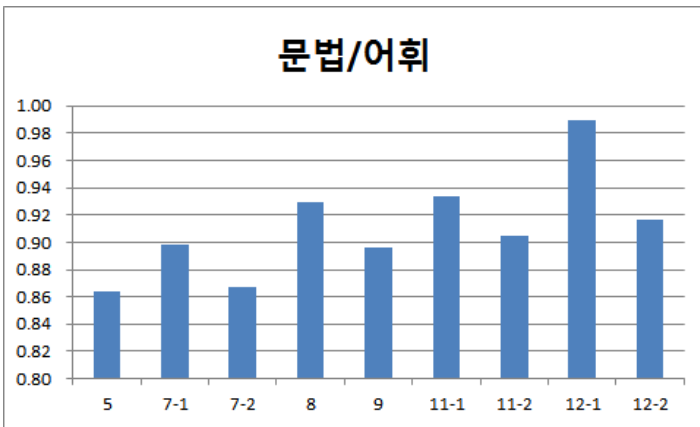
11) 특히 본 논문에서는 자료 전반에 대한 일반적이고 포괄적인 특성을 검토하고, 이를 바탕으

#### 4.1. 총량적 검토

먼저 각 학년별 자료에서 문법 형태소와 어휘 형태소의 비율을 정리하기로 한다. 각 형태소 가운데, 문법 형태소에는 모든 어미와 조사를 포함하되, 서술격조사와 접사는 모두 산입하지 않았다. 어휘 형태소에는 모든 단어를 포함하되, 어근은 포함하고 보조용언은 제외하였다.<sup>12)</sup> 이를 바탕으로 비율을 검토하면 다음과 같다.

〈표 4〉 자료별 어휘형태소에 대한 문법형태소의 비율

학년	5	7-1	7-2	8	9	11-1	11-2	12-1	12-2
비율	0.864	0.899	0.868	0.930	0.896	0.934	0.905	0.990	0.917



〈그림 1〉 자료별 어휘형태소에 대한 문법형태소의 비율

〈표 4〉와 〈그림 1〉에서 볼 수 있는 것처럼 학년이 올라갈수록 문법형태

로 중등학교 학생들의 작문 능력 판단의 변향으로 사용할 가능성을 점검하기로 한다.

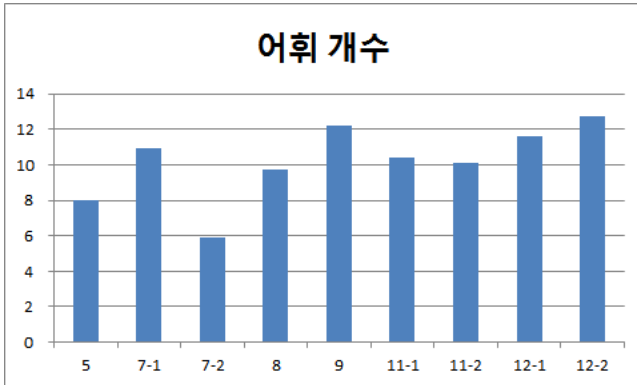
12) 보조용언은 어휘 사용 양상을 측정하는 데서는 양면성을 가진다. 형태론적으로는 어휘라고 할 수 있지만, 문법 범주의 표현 방식에서나 서술자의 개념 표현의 측면에서는 문법 범주를 나타내는 경우와 흡사하기 때문이다. 따라서 어휘 사용 능력을 측정하는 데서는 보조용언을 제외하고, 문장당 어절수를 계산하는 데서는 보조용언을 포함하기로 한다.

소가 차지하는 비중이 늘어남을 알 수 있다. 특히 8학년(중2)을 제외하면 어휘형태소에 대하여 문법형태소가 90%를 전후로 하여 사용되느냐를 기준으로 구분해 볼 수 있음을 알 수 있다.<sup>13)</sup>

이와 유사한 조사이지만 문장당 몇 개의 어절, 즉 몇 개의 어휘가 한 문장 안에 쓰였는지는 다음 <표 5>와 같다. 앞서와 달리 여기서는 보조용언을 어휘에 포함하여 계산하였다.

<표 5> 자료별 문장당 어휘의 개수

학년	5	7-1	7-2	8	9	11-1	11-2	12-1	12-2
개수	7.99	10.95	5.88	9.73	12.23	10.39	10.12	11.61	12.75



<그림 2> 자료별 문장당 어휘의 개수

<표 5>와 <그림 2>에서 보이는 사실 역시, 대개 12학년으로 갈수록 문장 안에 포함된 어절의 개수가 늘어남을 알 수 있다. <표 5>의 내용을 기초사된 문장의 길이, 문장을 구성하는 단어의 개수와 비교하여 보면, 물결 21에서 조사한 신문 기사의 평균 문장 길이가 대개 13개로 조사된 것을 보면(정유진 외 2013: 18), 학년이 늘수록 문장력이 성인의 그것과 유사해지는 경향을 보이고 있다고 할 수 있다.

다만, 어휘의 개수가 <표 4>에서 볼 수 있는 문법형태소의 개수와 비례관

13) 7-2는 선생님께 보낸 사신(私信)이고, 12-1은 주어진 질문에 답하는 방식의 자기소개서라는 점에서 다른 글과 장르상 차이가 있음에 유의하여 볼 필요가 있다.

계를 보이는 것은 아님을 알 수 있다.<sup>14)</sup> 예외가 없는 것은 아니지만, 우리는 학생의 발달에 비례하는 글쓰기 능력을 측정할 수 있는 사례는 문장을 이루는 어절의 개수보다는 문법형태소와 어휘형태소의 비율이 더 적극적으로 사용될 수 있을 것으로 추정한다.

## 4.2. 어휘 부류의 사용 양상

먼저 용언류와 체언류의 사용 양상에 대하여 살펴보기로 한다. 우리는 이 가운데 서술격조사 구문을 잠정적으로 제외하고,<sup>15)</sup> 동사문과 형용사문의 서술어가 명사류와 어떤 상대 빈도를 보이는지 점검하기로 한다. 이 상대 빈도는 다음과 같은 방식으로 구하였다.<sup>16)</sup>

(13) 동사문·형용사문의 서술어와 명사류의 상대 빈도

ㄱ. 상대값:  $(NNB+NNG+NNP+NP+NR+XR) - (XSA+XSV) - VCP$

ㄴ. 기준값:  $(VA+VV+VCN) + (XSA+XSV)$

ㄷ. 상대 빈도 = 상대값 / 기준값

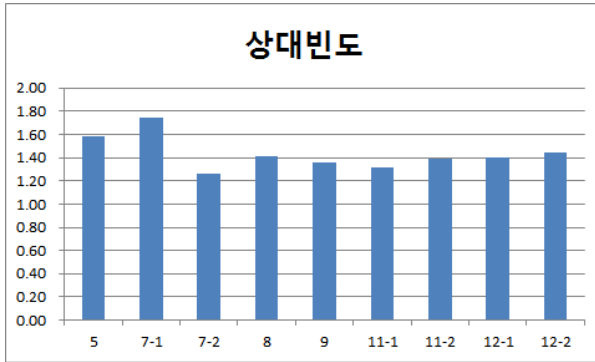
〈표 6〉 동사문·형용사문의 서술어와 명사류의 상대 빈도

학년	5	7-1	7-2	8	9	11-1	11-2	12-1	12-2
상대빈도	1.58	1.74	1.27	1.41	1.35	1.31	1.39	1.40	1.45

14) 보조용언을 계산에서 뺀 경우에도 결과는 <표 4>와 크게 달라지지 않았다. 따라서 이와 같은 차이가 보조용언을 산입하였느냐의 여부에 의한 것은 아니라고 하겠다.

15) 궁극적으로 서술격 조사에 선행하는 체언이나 어근이 갖는 속성에 따라서, 서술성을 가진 경우에는 동사나 형용사와 유사한 방식으로, 서술성을 가지지 않는 경우에는 명사와 유사한 방식으로 처리하여야 할 것이다.

16) 여기에서 제시하는 상대 빈도는 정확한 값이 아닌, 대략적인 수치를 보여준다. 부류를 단위로 해서 구분별 특성을 반영하기 어려운 경우가 발생하기 때문이다. 이를 테면, '-계'와 같은 부사파생접미사(XSB)의 경우에는 파생 부사의 어근이 문장성분을 취하기도 하고 취하지 않기도 하기 때문이다. 본 연구에서는 이들의 경우 (파생용언이든 근원어든) 용언의 뒤에 접사가 통합하는 경우로 분석하여 이와 같은 문제를 일관되지만 예외를 최소화하는 방안을 취하였다. 어근(XR)으로 처리된 경우는 '독특, 궁급, 솔직'처럼 대개 고립적인 비활성어근으로서 파생용언의 어간인 경우가 많았다는 경험적 기반에 근거하여 용언 범주로 넣되, 이들은 접사의 수효에서 모두 포함되는 것으로 계산하였다.

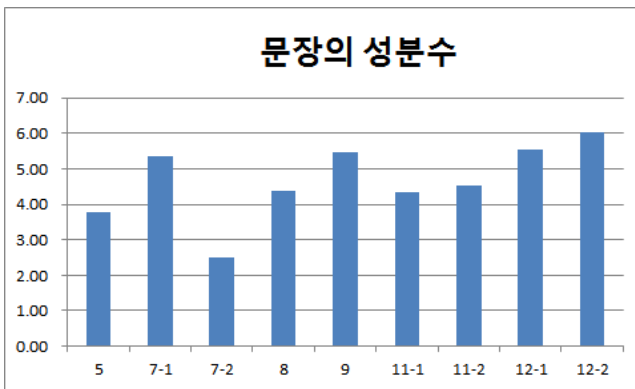


〈그림 3〉 서술어와 명사류의 상대 빈도

이와 유사한 방식의 비율을 검토하기 위하여 기준값을 종결어미(EF)로만 바꾸어서 검토하면 다음 <표 7>, <그림 4>와 같다.

〈표 7〉 문장당 체언류 성분의 상대 빈도

학년	5	7-1	7-2	8	9	11-1	11-2	12-1	12-2
상대빈도	3.77	5.37	2.49	4.39	5.46	4.34	4.52	5.54	6.03



〈그림 4〉 문장당 체언류 성분의 상대 빈도

<표 7>, <그림 4>의 전체 경향은 몇몇 유형(9, 11-1, 12-1)을 제외하고는 <표 5>, <그림 2>와 유사하다.

다음으로 어휘 부류별 고빈도 어휘를 살펴보기로 한다. 어휘의 계량이 ‘확률이 높고 낮은(more or less probable) 어휘 항목을 특정해 내는 작업이라고 할 때(김광해 1994), 빈도를 측정하는 양적 연구 방법은 크게 두 가지 방식에 의거할 수 있다. 그 하나는 개별 어휘의 빈도를 측정하는 방법이요, 나머지 하나는 어휘 총량의 분포를 살펴보는 방법이다. 전자의 연구 방법은 기껏해야 1000어절을 넘지 못하는 학생들의 낱말의 글을 대상으로는 유의미한 결과를 추출하는 작업을 하기가 매우 까다로울 것으로 추정된다. 본 연구에서는 따라서 총량의 분포에 주목하는 방안을 주로 적용하되, 그 안에서 고빈도를 보이는 어휘의 목록을 제시하고, 기왕의 조사 자료를 대조하여 보이기로 한다. 고빈도 어휘는 되도록 누적빈도 90%에 이르는 어휘까지를 보이는 것을 원칙으로 한다.

## 5. 결론을 대신하여

이 연구의 궁극적 목적에 값하기 위하여, 앞서 조사에서 시사하였던 몇 가지 경향에서 추적해 볼 만한 몇 가지 경향을 과제로 확인하여 두고자 한다. 다만, 특히 몇몇 자료(이를 테면, 5, 7-2, 11-2, 12-2 등)에서는 일관성을 의심할 만한 특징적 현상이 나타났음을 전제하여 두고자 한다.

(14) 언어 능력과의 상관관계를 의심하여 볼 사항

- ① 어휘 형태소에 비하여 문법 형태소의 비율이 증가해 가는가: <표 4, 그림 1>
- ② 문장당 어휘의 개수: <그림 2>
- ③ 서술문 명사류의 상대 빈도: <표 6, 그림 3>
- ④ 문장당 체언류 성분의 상대 빈도: <표 7, 그림 4>

## 6. 맺음말

본 연구에서는 전라북도 지역을 중심으로 하는 5학년, 7~9학년, 11~12학년 학생들이 작성한 글을 대상으로 계량적 방법으로 자료의 특성을 확인할 수 있는 계량적 방법론을 제안하였다. 특히 공개 프로그램을 바탕으로 실제 분석 과정에서 발생한 문제점들을 바탕으로 자료의 처리 방안과 분석 방법을 구체적으로 제시하였다. 이를 통해 본 연구의 분석 방법을 적용할 수 있는 객관적인 방법론으로 제시하였다.

언어 능력 변화를 추출하려는 시도로 접근한 것이지만, 글쓰기 자료에 대하여, 또한 본 조사에서 실시한 조사 결과에 대하여 평가는 되도록 언급하지 않으려고 하였다. 그 까닭은 크게 두 가지로 요약된다. 첫째, 본 조사는 ‘연구’의 단계로 탐구되지 않은 자료로서, 타당성에 대한 검증과 함께 추가적인 탐구가 필요하기 때문이다. 둘째, 글쓰기 자료에 대한 질적 평가와 함께 본 자료에서 보인 통계 결과가 평가되어야 하기 때문이다.

다만 기본분석된 자료를 대상으로 포괄적으로 추출된 몇 가지 검토 사항은 다음과 같이 변항으로 활용해 볼 가능성을 제시하고자 한다.

- ① 어휘 형태소에 비하여 문법 형태소의 비율이 증가해 가는가: <표 4, 그림 1>
- ② 문장당 어휘의 개수: <그림 2>
- ③ 서술문 명사류의 상대 빈도: <표 6, 그림 3>
- ④ 문장당 체언류 성분의 상대 빈도: <표 7, 그림 4>

본 연구는 구체적인 방법론을 바탕으로 중등학교 학생들이 쉽게 작성해 낼 만한 텍스트를 분석함으로써 언어 능력, 특히 작문 능력과의 상관관계를 추구하고 볼 추가적인 연구를 기다리고 있는 셈이다. 그럼에도 이와 같은 소규모 자료를 객관화하여 제시함으로써 이와 유사한 중등학교 학생들의 말, 글 자료를 비교해 볼 수 있는 기회를 제공할 수 있으리라고 판단한다. 본 연구의 방법론을 적용하여 구체적인 언어 사실을 어떠한 방식으로 확인하고 어떤 변항을 점검할 수 있을지는 추가적인 연구를 통하여 보고하고자 한다.

## 〈참고문헌〉

- 강범모(1999), 『한국어의 텍스트 장르와 언어 특성』, 고려대학교 출판부.
- 김광해(1993), 『국어 어휘론 개설』, 집문당.
- 김광해(1994), 『문체와 어휘』, 박갑수 편저, 『국어문체론』, 대한교과서 주식회사.
- 김유리(2015), 『1~2 학년군 초등 국어 교과서 분석을 통한 받아쓰기 교육 자료 개발 연구』, 전북대학교 교육대학원 석사학위논문.
- 김한샘(2005), 『현대 국어 사용 빈도 조사2』, 국립국어원.
- 김흥규·강범모(2000), 『한국어 형태소 및 어휘 사용 빈도의 분석1』, 고려대학교 민족문화연구원.
- 안예리(2013), 『20세기 전반기 국어의 문장 구성에 대한 연구』, 연세대학교 박사학위논문.
- 안예리·이주현(2014), 『20세기 문어 말뭉치 구축을 위한 기초 연구』, 『한국어학』63, 한국어학회, 229~265면.
- 이영제·강범모(2014), 『현대국어 역사 코퍼스를 이용한 언어 변화의 계량적 연구』, 『한국어학』63, 한국어학회, 267~303면.
- 李知英(2002), 『서사시제에 대하여-서사의 현재시제를 중심으로』, 朝鮮語研究會 編, 『朝鮮語研究 1』, くろしお出版, 147~166면.
- 정유진·김일환·강범모·김흥규(2013), 『신문의 언어 사용 통계』, 소명출판.
- 조남호(2002), 『현대 국어 사용 빈도 조사 -한국어 학습용 어휘 선정을 위한 기초 조사』, 국립국어연구원.
- 한송화(2013), 『한국어 접속부사의 사용 양상』, 『언어사실과 관점』31, 연세대학교 언어정보연구원, 139~169면.
- 황석자(1984), 『현대문체론』, 한신문화사.
- Biber, Douglas(2012), Register and discourse analysis. in James Paul Gee & Michael Handford eds, *The Routledge Handbook of Discourse Analysis*. London and New York: Routledge. pp. 191~208.
- Biber, Douglas & Susan Conrad(2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.

## 【국문초록】

이 연구는 중등학교 학생의 평가된 작문 말뭉치를 중심으로 언어 특성을 추출하여 비교함으로써 텍스트를 바탕으로 계량적 평가 가능성이 있는 항목을 추출하여 제안하려는 데 목적이 있다. 그리하여 전라북도 지역의 중등학교 학생들이 실제로 작성한 글을 대상으로, 분석 과정을 실질적으로 제시하고, 이를 현장에서 사용할 수 있는 공개적이고 계량적인 방안으로 계발하고자 하였다.

본 연구에서 제안한 방법론은 학교 현장에서 실제 적용이 가능할 수 있도록 구안하였다. 따라서 공개 프로그램을 바탕으로 실제 분석 과정에서 발생한 문제점들을 바탕으로 자료의 처리 방안과 분석 방법을 구체적으로 제시하였다. 이를 통해 본 연구의 분석 방법이 좀 더 객관적으로 적용될 수 있는 가능성을 탐색하였다.

언어 능력 변향을 추출하려는 시도로 접근한 것이지만, 글쓰기 자료에 대하여, 또한 본 조사에서 실시한 조사 결과에 대하여 평가는 되도록 언급하지 않으려고 하였다. 그 까닭은 본 조사 연구는 글쓰기 자료에 대한 질적 평가와 함께 본격적으로 검증을 기다리는 매개변인 탐색에 치중하였기 때문이다.

본 연구 조사를 통해서 포괄적으로 추출된 몇 가지 검토 사항은 다음과 같이 변향으로 활용해 볼 가능성을 제시하였다.

- ① 어휘 형태소에 비하여 문법 형태소의 비율이 증가해 가는가
- ② 문장당 어휘의 개수
- ③ 서술문 명사류의 상대 빈도
- ④ 문장당 체언류 성분의 상대 빈도

주제어: 글쓰기, 평가 변향, 중등학교, 말뭉치 언어학, 형태소 분석, 품사, 어휘 형태소, 문법 형태소

【Abstracts】

## Methods for assessing the writing ability of middle school students

Seo Hyeong-guk

This paper aims at suggesting the computer- and corpus-based methods for assessing writing ability of middle school students. For the purpose of this survey, I executed:

- 1) collecting middle school student's writing materials.
- 2) parsing those materials on the basis of 21 Sejong Project tagset.
- 3) analysing parsed corpus for the grammatical properties.

This paper suggest only the large-scale parameters for assessing writing ability, and the rest, i.e. every small categorial parameters are remained for later.

To sum up, this paper suggests parameters as:

- 1) grammatical morpheme ratio
- 2) lexemes per sentence
- 3) nouns per predicate
- 4) nouns per final ending

Key words: assessment, writing ability, middle school students, corpus linguistics, parsing, Part-Of-Speech, noun, sentence, verb, adjective, usage

이 논문은 2015년 7월 6일에 투고되었으며, 2015년 8월 5일에 심사 완료되어 2015년 8월 12일에 게재가 확정되었음.

