

초등 읽기 어휘 조사 도구의 설계

배진성^o, 김갑수

서울신목초등학교, 서울교육대학교 컴퓨터교육과
bjs114@unitel.co.kr, kskim@ns.seoul-e.ac.kr

Design of Investigating Tool of the Vocabulary of the Korean Language Reading Textbooks for the Elementry School

Jin-Seong Bae^o, Kab-Su Kim

Seoul Sinmuk Elementary School, Dept. of Computer Education, Seoul National University of Education

요 약

초등학생들에게 읽기 교과서는 어휘 습득에 있어 의도적인 자료의 시작으로, 그 어휘를 조사, 계량화하는 일의 의미는 매우 크다고 할 수 있다. 어휘 조사 과정의 중요한 작업은 많은 양의 자료를 색인, 계량하는 일이다. 이러한 자료 처리에 컴퓨터를 이용하면, 많은 부분을 자동화 할 수 있으며, 분석된 자료를 이용하여 다양한 작업이 가능할 것이다.

많은 양의 자료를 원활하게 처리하기 위해서는 적절한 도구가 필수적이다. 이에 어휘 조사 과정을 보다 편리하게 하고, 자동으로 처리할 수 있으며, 만들어진 자료를 재활용 할 수 있는 어휘 조사 도구를 설계하게 되었다.

1. 서 론

초등학생들에게 있어, 읽기 교과서는 어휘 습득에 관한 의도된 자료로서, 사용된 어휘의 중요성은 두 말할 필요가 없을 것이다. 이러한 어휘를 조사하여 계량화하는 것은 매우 중요한 일이다.[1]

어휘의 양적 구조와 체계 및 운용 등을 밝히기 위해서는 각종 자료를 통한 어휘의 조사가 선결 과제이다. 어휘의 조사 없이 이들 문제를 파악한다는 것은 불가능에 가까운 것으로, 그 동안 어휘 조사가 거의 이루어지지 않은 우리에게는 매우 시급한 과제 중의 하나이다[2].

어휘를 조사하는 과정이 복잡하지는 않으나 그 양이 많고, 어휘의 빈도, 색인 등 다양한 작업을 일일이 손으로 하기에는 많은 시간이 걸린다. 이에 어휘 조사의 과정을 자동으로 처리할 수 있는 도구의 개발이 필수적이다.

또한 대부분의 어휘의 조사 연구들을 살펴보면 일차적인 현황에 치중되어 있으며, 또한 조사 결과가 대부분 문서 형태로 보관되어 재활용의 어려움이 따른다.

이에 본 연구에서는 초등학교 교과서 중에서 가장 어휘 조사와 관계가 깊은 읽기 교과서를 선택하여, 어휘 분석을 할 수 있는 도구를 설계하게 되었다.

따라서 본 연구의 목적은 다음과 같다.

첫째, 어휘 분석을 자동화 할 수 도구를 설계한다. 이를 통해 보다 빠르고, 편리하게 어휘 분석 작업을 할 수 있을 것이다.

둘째, 어휘 분석의 결과들은 재활용 가능하도록 DB의 형태로 제공할 수 있도록 설계한다. DB로 저장하게 되면 컴퓨터를 이용하여 많은 양의 자료를 쉽게 조작할 수 있으며, 통계적 작업을 위한 다양한 결과를 산출할 수 있을 것이다.

2. 이론적 배경

2.1 어휘 조사

어휘는 어의 그대로 어의 집합이다. 일반적으로 일정한 범위에서 사용되어 형성되는 어의 총체를 말한다. 이때, 그 어의 총체는 어떠한 언어체계로서 어떠한 관점에서 선택되어져야 한다.

어휘단위는 단어와 일치되지 않아 단어보다 작은 형태소, 단어보다 큰 관용어를 포함한다. 특히 관용어는 문법적으로 규정하기 어려운 것으로 어떤 언어 집단에서는 통용되지 않으므로 그것은 그 언어집단의 문화양식과 관계가 있다[4].

어휘 조사에 있어서 그 목적에 따라 일정한 기준으로 설정한 뒤 그 계량의 기준이 되는 단위, 즉 헤아려 나가는 단위를 조사 단위라고 한다. 조사 단위를 설정하는 것에 따라 어휘 조사의 결과는 커다란 차이를 나타낼 수 있다[2]

특히 국어는 용언의 변화형이 매우 복잡할 뿐만 아니라 합성 동사와 보조 용언의 쓰임도 매우 다양하며, 명사에 있어서도 합성 명사와 그 조사의 쓰임이 매우 복잡하여, 어휘 조사에 있어서 신중을 기해 조사 단위를 정해야 할 것이다.[5]

어휘를 대상으로 하는 조사 작업은 언어학적 지식과 통계학적 지식이 결합되어 이루어진다.

2.2 형태소 분석과 형태소 분석기

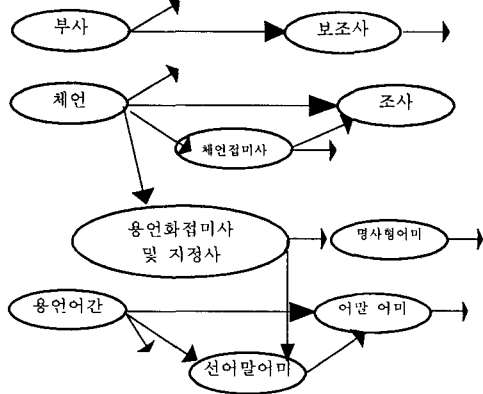
형태소는 일정한 음성에 일정한 뜻이 결합되어 있는 말의 가장 작은 단위이다. 즉, 최소 유의적 단위(minimum meaning unit)라고 정의되며, 어소(morpheme)라고도 한다. 이는 형태론의 최소 단위가 되는 동시에 음운론의 상위 단위가 되고, 어절, 구, 문장 등을 구현하는 요소이다[5,6]

한국어는 교착어로서 문장의 단위로 할 수 있는 어절

이 어휘 형태소와 문법 형태소의 결합으로 이루어져 있으며, 문법 형태소가 문장에서의 문법적 기능을 지시하는 특징을 가지고 있다. 이는 굴절어라 할 수 있는 영어나, 독일어 등의 서구 언어와 명백한 차이를 보이는 부분이기도 하다.

형태소 분석이란 자연어의 입력 문자열로부터 문장의 해석에 필요한 단위 문자열로 잘라서 각종 지식 정보, 즉 품사의 범주, 활용형 등 각종 정보를 추출해 내는 과정을 형태소 분석이라고 한다.[7]

한국어는 어절간의 띄어쓰기가 있으므로 어절 단위로 분석이 가능하다. 따라서 형태소 분석에서는 어절 내부에서 형태소간의 결합 및 상태 전이에 따라 분석이 이루어지는데, 개략적인 모델은 다음과 같다.[8]



< 그림 1 > 어절내부의 형태소 결합방

형태소 분석기(morphological analyzer)는 한국어 처리를 위해서 하나의 어절이 가지는 형태소를 추출하고, 이 형태소가 가지는 정보(품사, 원형, 의미 등)가 무엇인가를 찾아내는 시스템을 말한다[9].

한국어 형태소 분석 기법으로는 한국어의 단어 형성 규칙에 따라 여러 가지 다양한 방법론이 제시되어 왔으며, 형태소간 결합 규칙을 이용하여 어절을 형태소로 분리하는 것이 효율적이다. 일반적으로 많이 알려진 것으로는 Head-Tail 구분법, Tabular 파싱법, 최장일치법과 최단일치법, 음절 단위 분석법, 사전-중심 형태소 분석법 등 다양한 모델들이 제안되었다.

3. 어휘 조사 도구

3.1 어휘 조사 도구의 시스템 구조도

교과서를 대상으로 한 어휘 조사 연구는 많지 않으나 안승덕, 김재운(1975)의 "국민학교 국어 교과서의 어휘 연구"[2]와 박봉배(1975)의 "초등교육에 있어 우리말 기본 학습 어휘에 관한 연구"[10]를 비교 분석하였다.

안승덕, 김재운(1975)의 "국민학교 국어 교과서의 어휘 연구"는 전학년 읽기 교과서를 대상으로 하여, 국어 교과서에 나오는 총어휘의 수와 학년별 어휘의 수를 비교할 수 있었으며, 박봉배(1975)의 "초등교육에 있어

우리말 기본 학습 어휘에 관한 연구"에서는 그 대상이 1, 2학년에 한정되어 있으나, 학년별로 타 교과서와의 차이를 구별할 수 있었다.

위의 연구에서 어휘 조사에 필요한 기본적인 모듈이 갖춰야 될 기능을 추출할 수 있었는데, 총어휘의 수, 어휘의 반복정도, 어휘의 품사별 분류, 학년별 어휘의 수의 비교 등 필요한 기능들을 추출할 수 있었다.

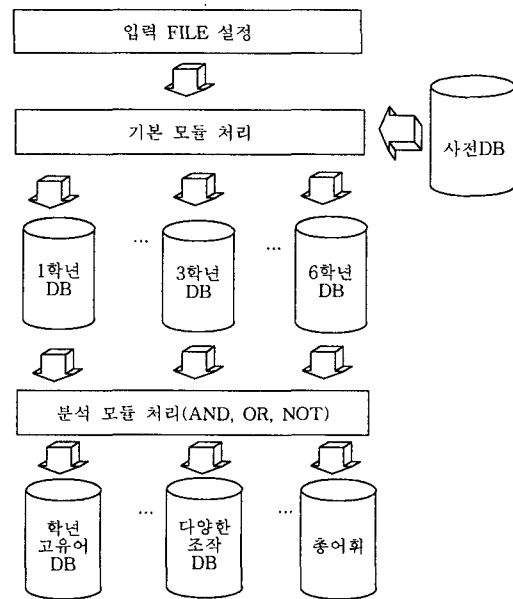
본 어휘 분석 도구의 특징은 다음과 같다.

첫째, 많은 양의 자료를 처리하므로, 모든 결과물은 DB로 만들어지도록 했다. DB로 만들면 비교, 검색 등 다양한 자료 처리 작업이 쉬우며, 다양하게 활용될 수 있을 것이다.

둘째, 다양한 계량적 분석을 위해 몇 개의 모듈의 통합 형태로 이루도록 설계하였다. 이를 통해 중간에 만들어지는 자료DB의 재활용을 꾀하였다.

셋째, 특정한 자료 DB간의 상호 분석이 가능하도록 설계하였다. 이를 통해 학년별 어휘를 추출한 후, 그 비교를 쉽게 할 수 있을 것이다.

넷째, 사전-중심 형태소 분석법(dictionary-oriented)을 이용한다. 사전-중심 형태소 분석법은 형태소 분리가 형태론적 변형 문제를 사전을 중심으로 하여 처리하는 기법이다. 사전 중심의 처리 방법은 알고리즘으로 처리하는 비중을 줄이고 사전의 비중을 크게 하여 형태소 분석 과정의 일부를 사전에 의존한다[9].



<그림 2> 어휘 조사 도구의 시스템 구조도

3.2 설계

1) 입력 파일

어휘 조사 도구에서 사용될 자료의 형식을 다음과 같이 설계했다.

㉠ ks5601완성형 코드로 작성되며, 즉, 메모장과 같은 간단한 텍스트 입력기에서 입력과 편집이 가능하다.

“훈글”에서는 본문을 입력한 후, “새 이름으로” 저장하여 쉽게 만들 수 있다.

㉡ 자료의 시작부분에 쪽번호를 #1, #2, #123 과 같이 쪽수를 입력하도록 했다.

㉢ 지시문 앞에는 ‘*’표시를 하도록 했다. 지시문은 본문의 내용과 달리, 같은 말이 계속 반복되며, 어휘 조사 경우 그 빈도가 매우 많으므로 무시할 수 있도록 하기 위해서, 특정 표시를 달게 하였다.

```
#8 <- 쪽번호
*글을 읽고, 글의 내용을 간추려 봅시다. <- 지시문
*언제, 어디에서, 무엇을 하였는지 생각하며 '유관순'을 읽어 봅시다. <- 지시문
유관순
유관순은 1902년 3월 15일, 지금의 충청남도 천안시 병천면 지령 마을에서 태어났다.
.....
.....
```

<그림 3> 입력 형식 예

2) 기본 모듈

기본 모듈에서는 교과서에 등장하는 모든 어휘를 조사하여, 가나다순으로 등록하여 DB를 만들어야 한다. 아울러 추출된 어휘의 빈도수를 세어 기록한다.

또한 어휘의 품사를 결정하여, DB에 기록한다. 품사는 크게 명사, 대명사, 수사, 동사, 형용사, 관형사, 부사, 감탄사의 8대 품사로 나누게 된다.

3) AND 분석 모듈

기본 모듈을 이용하여 처리된 두 개의 DB를 분석하여, 서로 겹치는 낱말을 추출하여 새로운 DB를 생성하는 모듈이다.

예를 들면, 기본 모듈을 이용하여 1, 2 학년의 어휘를 각각 처리한 후, AND 분석 모듈을 이용하면 1학년과 2학년의 공통어를 추출할 수 있게 된다. 즉 AND 분석 모듈을 이용하면 손쉽게 공통어를 얻을 수 있다.

4) OR 분석 모듈

기본 모듈을 이용하여 처리된 두 개의 DB를 분석하여, 두 기본 모듈의 합을 생성하는 모듈이다. 이를 이용하면, 교과서에 등장하는 모든 낱말을 구할 수 있다.

5) NOT 분석 모듈

기본 모듈을 이용하여 처리된 두 개의 DB를 분석하여, 겹치는 낱말을 제외한 나머지 어휘를 DB로 생성하는 모듈이다. 이를 이용하면, 특정 학년의 고유어를 추출할 수 있다.

6) 출력 파일

㉠ 기본 모듈 DB 테이블

column_name	data_type	설 명
Index	integer	색인번호
Word	varchar	어휘(단어)
Word_count	integer	어휘의 빈도수
Word_part	integer	어휘의 품사

㉡ 분석(AND, OR, NOT) 모듈 DB 테이블

column_name	data_type	설 명
Index	integer	색인번호
Word	varchar	어휘(단어)
Word_part	integer	어휘의 품사

4. 결론 및 향후 과제

흔히, 국어 교과를 도구 교과라 하며, 이는 다른 교과 학습에 지대한 영향을 미치는 것을 의미한다. 즉, 타 교과의 학습 성취와 깊은 관련이 있음을 뜻하며, 올바른 어휘에 대한 이해 없이 좋은 학습 성취를 거두기는 매우 어렵다.

학년간의 어휘의 차이를 이해하고, 발달과정에 맞는 어휘의 선별, 이에 맞는 교과서의 집필과 교육에 대한 끊임없는 연구는 필수적이다.

본 논문에서는 초등학교에서 읽기 교과서를 중심으로 그 어휘를 조사할 수 있는 도구를 설계해 보았다.

어휘 조사에 있어 주된 작업은 많은 양의 자료를 처리하는 일이며, 이에 자동화된 도구의 사용은 필수적이다.

향후 본 논문에서 제시된 도구를 구현함으로써, 어휘 조사에 있어 자동화된 도구의 효용성을 규명하고, 다양한 학년별 자료DB를 생성하여 타 교과에 적용, 확장해 볼 것이다.

참고 문헌

- [1] 안승덕, 김재윤, “국민학교 국어 교과서의 어휘 조사 연구”, 청주교육대학 논문집, 제11집, pp125-181, 1975
- [2] 김광해, “국어 어휘론 개설”, 집문당, 1993
- [3] 이태욱, “컴퓨터 교육론”, 좋은 소프트, 1999
- [4] 임만영, “국민학교에서의 어휘교육”, 국어과교육연구, Vol.6 No.1, 1988
- [5] 서정수, ‘국어 문법’, 뿌리깊은나무, 1984
- [6] 남기심, 고영근, ‘표준 국어 문법론’, 탑출판사, 1987
- [7] 박기홍, 이태현, 정민수, “품사 사전을 이용한 형태소 분석”, 군산대학교 논문집, Vol.23 No.1, 1996
- [8] 서상규, 한영균, “국어정보학 입문”, 태학사, 1999
- [9] 이용재, “자연어를 이용한 관광안내 시스템을 위한 형태소 분석기의 개발에 관한 연구”, 자연과학논문집, Vol.10, No.1, 1998
- [10] 박봉배, “초등교육에 있어 우리말 기본학습 어휘에 관한 조사연구”, 서울교육대학교 논문집, Vol.8 No.1, 1975