

## Korean Learning Assistant System with Automatically Extracted Knowledge

Gi-Tae Park<sup>†</sup> · Tae-Hoon Lee<sup>†</sup> · So-Hyun Hwang<sup>†</sup> · Byeong Man Kim<sup>\*\*</sup>  
Hyun Ah Lee<sup>\*\*\*</sup> · Yoon Sik Shin<sup>\*\*</sup>

### ABSTRACT

Computer aided language learning has become popular. But the level of automation of constructing a Korean learning assistant system is not so high because a practical language learning system needs large scale knowledge resources, which is very hard to acquire. In this paper, we propose a Korean learning assistant system that utilizes easily obtainable knowledge resources like a corpus, web documents and a lexicon. Our system has three modules - problem solving, pronunciation marker and writing assistant. Automatic problem generator uses a corpus and a lexicon to make problems with one correct answer and three distracters, then verifies their suitability by utilizing frequency information from web documents. We analyze pronunciation rules for a pronunciation marker and recommend appropriate words and sentences in real-time by using data extracted from a corpus. In experiment, we evaluate 400 automatically generated problems, which show 89.9% problem suitability and 64.9% example suitability.

**Keywords :** Korean Learning Assistant, Automatic Problem Generation, Automatic Korean Pronunciation Marking, Automatic Writing Assistant

## 자동 추출된 지식에 기반한 한국어 학습 지원 시스템

박 기 태<sup>†</sup> · 이 태 훈<sup>†</sup> · 황 소 현<sup>†</sup> · 김 병 만<sup>\*\*</sup> · 이 현 아<sup>\*\*\*</sup> · 신 윤 식<sup>\*\*</sup>

### 요 약

정보통신기술을 활용한 학습시스템은 꾸준히 강조되어 왔지만 한국어 학습시스템의 자동화 수준은 높지 않다. 실용성 있는 학습시스템의 구축에는 대량의 기반지식이 필요하지만 이러한 지식을 구축하기 쉽지 않기 때문이다. 본 논문에서는 한국어학습시스템의 요소로 어학문제풀이, 표준발음 도우미, 글쓰기 도우미를 제안하고, 획득이 용이한 말뭉치와 웹문서, 사진을 활용하여 구축된 학습지원시스템을 소개한다. 어학문제풀이를 위한 자동문제생성에서는 말뭉치와 사진을 이용하여 문제와 보기문항을 생성하고, 웹문서 검색빈도를 활용하여 보기적합성을 검증한다. 표준발음 변환을 위해서 발음표기법을 분석하였으며, 글쓰기 지원을 위해 말뭉치에서 추출한 기본식데이터를 이용한 실시간 어휘추천과 문장추천을 구현하였다. 실험에서는 제안하는 방법으로 생성된 임의의 400문제에 대한 판정 결과 89.9%의 문제 적합률과 64.9%의 보기 적합률을 보였다.

**키워드 :** 한국어학습시스템, 자동문제생성, 자동발음표기, 자동글쓰기지원

### 1. 서 론

21세기 국제화 시대에 한국은 기술적 문화적 우수성을 인정받으며 그 위상이 날로 높아지고 있다. 지리적으로 인접한 아시아뿐만 아니라 범 유럽에서도 한국어를 배우려는 학습

열풍이 대단하다. Fig. 1에서 볼 수 있듯이 외국인을 대상으로 하는 한국어능력평가시험인 TOPIK의 응시자도 매년 급격히 증가하고 있다. 이러한 추세에 힘입어 우리 정부에서는 한국어 세계화를 추진하기 위해 부단히 노력하고 있으며, 최근에는 '2012년 한국어 세계화 3대 추진 과제'를 발표하는 등 한국어를 전파하기 위해 적극적인 모습을 보이고 있다.

한국어를 보다 쉽고 편리하게 배우기 위해서는 정보통신기술을 활용한 교수법이 필수적이다. 외국인들의 경우 한국어를 배우기 위해 현지에서 어학당과 같은 교육기관을 찾고 있지만 학습자의 가파른 증가세를 수용하지 못하고 있다. 현지인의 표현이 교육에 빠르게 반영되고, 시간

※ 본 연구는 금오공과대학교 학술연구비에 의하여 연구된 논문임.  
<sup>†</sup> 준 회 원 : 금오공과대학교 컴퓨터소프트웨어공학과 학부과정  
<sup>\*\*</sup> 정 회 원 : 금오공과대학교 컴퓨터소프트웨어공학과 교수  
<sup>\*\*\*</sup> 중신회원 : 금오공과대학교 컴퓨터소프트웨어공학과 부교수  
 논문접수 : 2012년 5월 9일  
 수정일 : 1차 2012년 7월 9일, 2차 2012년 7월 20일  
 심사완료 : 2012년 8월 9일  
 \* Corresponding Author : Yoon Sik Shin(ysshin@kumoh.ac.kr)

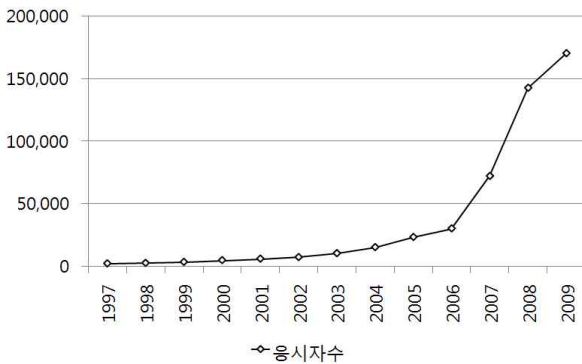


Fig. 1. TOPIK applicant fluctuation

과 장소, 수용인원에 구애받지 않는 교육을 위해서는 인터넷 학습 서비스가 적합하다. 이처럼 언어교육을 위해 정보통신기술을 활용하는 방법을 CALL(Computer Assisted Language Learning)이라고 부른다.

한국어능력이란 한국어의 올바른 사용능력을 말한다. 언어 능력은 말하기, 듣기, 읽기, 쓰기의 네 가지 영역으로 나뉘며 어휘 능력은 모든 영역의 공통적 기반지식이 된다[1]. 언어 능력 개발을 위한 학습 방식은 원칙을 먼저 배워 그 원칙에 따라 문장을 만들어내는 연역적 학습방법과, 수많은 예문을 접하여 저절로 문장구성의 원리를 깨달아가는 귀납적 학습방법으로 나뉜다[2]. 연역적 학습은 지도하기가 간편하며 어휘나 문법을 제시하는 방식으로 학습시스템을 구축하기 편리하지만, 자연스러운 언어 습득 과정인 귀납적 학습이 동반되지 않으면 학습의 지속성을 보장하기 어렵다. 귀납적 학습에서는 텍스트나 멀티미디어로 다양한 단어나 문장, 상황을 제시하거나 TOPIK과 같은 사지 선다형 문제를 통해 푸는 방식을 사용한다. 이러한 귀납적 학습을 위한 시스템에는 많은 양의 언어 지식이 필요하지만, 양질의 지식을 구축하기 위해서는 높은 비용이 요구되는 문제점이 있다[3].

본 논문에서는 비교적 획득하기 쉬운 말뭉치와 웹문서, 사전을 귀납적 언어학습을 위한 지식원으로 사용하여 자동으로 추출한 지식에 기반한 한국어 학습지원시스템을 제안한다. 시스템에서는 말하기, 듣기, 읽기, 쓰기의 언어 영역 중 텍스트에 기반한 읽기와 쓰기를 대상으로 하여, 한국어 어휘문제생성, 표준발음 표기변환, 글쓰기 학습을 지원한다. 사지선다형 어휘문제생성에서는 말뭉치에서 추출한 문장에 대해서 사전과 웹문서를 활용하여 보기를 생성하고 보기의 적합성을 검증한다. 표준 발음 표기에서는 표준 발음 표기 규칙의 적용 순서를 정의하여 표기의 정확성을 높인다. 글쓰기 학습지원에서는 말뭉치에서 추출한 정보를 활용하여 추천 조사와 용언, 문장을 실시간으로 제시한다.

## 2. 기존 연구

국내의 학습시스템의 자동화는 영어 학습에 집중되어 있다[4-7]. 한국어 학습을 위한 다양한 시스템이 구축되어 있

으나 대부분 수동으로 구축된 지식에 의존하고 있으며[3], 자동화 기술이 적용된 사례로는 과학기술부 및 한국과학재단 국가지정연구실 과제로 수행된 우리말 배움터[8]와 한국어 어휘학습시스템[9]이 있다. 우리말 배움터는 자동화된 맞춤법과 문법 검사기, 표준발음 변환기를 제공하고 있으나 수동으로 구축된 지식에 의존적인 부분이 많다. 한국어 어휘학습시스템은 사전과 WordNet과 같은 기반지식과 사용자 어휘지능망을 통해 9가지 유형에 대한 수준별 문제를 생성하는데, 다량의 기반지식이 필요하며 변화하는 언어 현상을 반영하는데 어려움이 있다.

한국어의 발음은 국립국어원에서 제시하는 표준발음법[10]이 기준이 된다. 표준발음법은 사람이 이해하기 쉽도록 원칙, 음운체계, 음운 변동으로 구성되며 총 7장 30항으로 구성되며 서로 간에 영향을 줄 수 있다. 표준발음법을 교육학적으로 이해하고 적용하기 위한 연구는 진행되었지만[11], 음운의 변동규칙에 따른 정보 유실을 방지하면서 소프트웨어적으로 자동화하기 적합한 절차에 대한 연구는 부족하다. 외국인의 한국어 발음 학습을 위한 연구[12]에서는 문장 내에서 발화된 어휘들의 정확한 발음을 익히기 위해서 사전식 발음학습이 아닌 임의의 문장에 대한 발음표기가 필요함을 지적하였다. 이에 기반하여 우리말 배움터에서는 임의의 문장에 대한 표준발음 변환기를 제공한다. 입력된 문장에 대하여 어휘망과 공기 관계(co-occurrence) 등을 활용하여 음운구를 설정하고 동형이의어의 중의성을 해소하여 높은 신뢰도를 갖는 변환을 제공하는 한국어 학습을 지원하고 있으나, 과도한 음운구 설정에 의한 문제와 트랩 오류 등의 예외에 취약한 문제가 있다.

학습요소 중 쓰기학습은 전체 학습요소 중 가장 취약한 부분이다. 외국인의 TOPIK 문제 풀이와 한국어 작문에 대한 연구 결과에서는 체언과 용언의 활용형, 조사와 같은 학습요소에 큰 어려움을 느끼는 것으로 분석되었다[13]. 한국어 작문에 대한 지원 시스템은 문법 검사기에 치중되어 있어 “밥을 먹는다”와 같은 문법적인 오류는 처리할 수 있으나 “밥이 먹는다”와 같은 의미적인 오류는 처리하지 못하고 있다. 외국인을 위한 대부분의 한국어 학습시스템은 학습자에게 제공된 문장을 따라 쓰게 하거나 빈칸에 문장을 기입하여 보는 방식으로 제공하고 있어 반복적이며 비효율적인 쓰기 학습 방식을 취하고 있다[3]. 한국인을 대상으로 하는 작문학습시스템에 관해 연구된 바 있지만[14, 15] 자동화를 기대하기 어렵고 외국인을 대상으로 하는 한국어 작문 시스템의 경우 관련연구도 부족한 상황이다.

## 3. 한국어 학습시스템

본 논문에서는 말뭉치와 웹문서, 사전에서 자동으로 추출한 지식에 기반한 한국어 학습시스템의 구축 방안을 제안한다. 논문에서 제안하는 시스템은 한국어에 대한 기본적인 학습이 이루어져 있는 초급 수준의 외국인을 가정한다. 아래에서는 제안하는 학습시스템의 어휘문제 자동생성, 자동 발음 표기, 글쓰기 학습지원에 대해서 설명한다.

3.1 어휘문제 자동 생성

사지선다 형태의 어휘문제는 TOEIC, TOEFL, TOPIK 등의 대부분의 어학 능력 평가에 사용되는 형태로 언어 학습의 기본인 어휘 학습의 수준을 평가하는데 이용된다. 사지선다형 문제는 빈칸이나 밑줄 친 어휘에 관련된 문제로 구성되어 어떤 어휘를 문제의 대상으로 삼는가와 함께 적합한 보기문항을 만드는 것이 출제의 주안점이 된다. 본 절에서는 한국어 어휘문제를 자동으로 생성하기 위해 TOPIK의 문제유형을 분석하고 선택된 유형의 문제를 자동으로 생성하는 방법과 문제생성의 검증을 위한 방법을 제안한다.

1) 어휘문제 유형 분석

본 논문에서는 문제생성을 위해 한국어능력시험인 국립국제교육원의 TOPIK을 활용한다. 문제의 출제도와 보기의 선정기준을 파악하기 위해 2003년부터 2010년까지의 기출문제 중 일상의 한국어능력을 평가하는 초급과 중급 문제를 분석하여 문제의 유형과 품사에 따라 대중소 분류로 구분한 Table 1의 분류를 얻었다. 문제는 크게 빈 칸에 알맞은 보기를 선정하는 빈칸 채우기 유형과 밑줄 친 어휘와 연관되는 보기를 선정하는 연관어휘 유형으로 나뉜다. 아래에서는 각 문제유형에 대해 설명한다.

Table 1. Class of TOPIK word problems

대분류	중분류	소분류
빈칸 채우기	적합 어휘 선정	T1. 적합 체언
		T2. 적합 용언
		T3. 적합 조사
		T4. 적합 부사
	용언 활용	T5. 다의적 용언
		T6. 어미 이해
연관 어휘	유의 어휘 판단	T7. 유의 체언
		T8. 유의 용언
		T9. 유의 부사
	반의 어휘 판단	T10. 반의 용언

a) 적합 어휘 선정 유형

문장의 빈칸에 적합한 어휘선택을 통해 어휘 이해도를 측정하는 문제유형이다. 어휘 능력을 판단하는 기본적인 형태로, 오답보기는 문맥상에서 의미적으로는 맞지 않지만 인접한 체언이나 조사, 용언과는 사용이 가능한 어휘로 구성된다.

b) 용언 활용 유형

용언을 괄호로 대체한 여러 문장에서 공통적으로 사용될 수 있는 다의어를 선택하는 다의적 용언 문제와, 단일 문장이지만 어미에 대한 이해능력을 확인하는 용언의 활용능력 문제를 포함한다. 다의적 용언 문제의 오답보기는 주어진 문장들 중의 일부에만 들어갈 수 있는 용언이 제시된다. 어미 이해 문제는 시제가 일치하지 않거나 문맥상 어울리지 않는 어미 조합을 오답보기로 제시한다.

c) 유의 어휘 판단 유형

문장 내에서 밑줄 친 어휘에 대한 유사어를 선택하는 문제로, 주어진 문맥에서의 어휘 의미의 올바른 이해와 유의어에 대한 이해능력을 평가한다. 오답보기는 동음이의어의 유의어, 반의어, 의미적 유의성은 없지만 문맥에 나타날 수 있는 어휘 등으로 구성된다.

d) 반의 어휘 판단 유형

문장 내 밑줄 친 어휘에 대한 반대되는 보기를 선택하는 문제로, 주어진 문맥에서의 어휘 의미의 올바른 이해와 반의어에 대한 이해능력을 평가한다. 주어진 오답보기는 유의어휘와 동일하게 동음이의어의 반의어, 유의어, 발생 가능한 반의어가 아닌 어휘 등으로 구성된다.

2) 말뭉치와 사전을 이용한 어휘문제생성

문제생성에서는 Table 1의 문제유형들 중에서는 T1 적합 체언, T2 적합 용언, T3 적합 조사, T8 유의 용언, T10 반의 용언을 대상으로 한정한다. 부사에 관련된 T4와 T9는 용언 문제생성방법을, 유의 체언에 대한 T7은 적합 체언에 대한 문제생성방법을 그대로 적용할 수 있다. 자동문제생성이 쉽지 않은 T5와 T6는 추후 연구과제로 둔다.

본 논문에서는 Fig. 2와 같이 형태소 분석, 문제 및 정답 생성, 보기 생성의 단계를 거쳐 문제를 생성한다. 사지선다형 문제생성에서 가장 어려운 점은 문제 대상의 선택과 적합성이 보장된 보기문항의 생성에 있다. 보기의 적합성이란 오답보기가 적절한 난이도를 보장하면서 두 개 이상의 정답이 나올 수 없어야 함을 의미한다. 문제 대상의 선택과 난이도 보장 방식은 본 절에서 설명하며, 하나의 정답을 보장하기 위한 문제 검증 방식은 다음 절에서 설명한다.

a) 문장 수집 및 형태소분석

세종 말뭉치나 교과서 문장과 같이 정갈한 원시 말뭉치의 문장을 모아 형태소 분석기를 이용해 태깅(tagging)한다. 형태소 분석 결과의 품사 태그는 유형별 문제유형 생성과 검증에서 활용한다.

b) 문제 및 정답 생성

두 번째 단계에서는 품사태깅결과에 기반하여 문장에서 발생하는 체언과 용언, 조사에 대해 가능한 모든 유형의 문제를 생성한다.

문제유형 중 T1, T2, T3는 빈칸에 적합한 어휘를 찾는 문제이다. 해당 유형의 문제와 정답보기는 문장에서 발생하는 체언, 용언, 조사에 대해서 원 문장에 있던 어휘를 정답보기로 선정하고 괄호로 치환하여 생성한다.

T8 유사 용언과 T10 반의 용언은 밑줄 친 용언의 유의어나 반의어를 이용하여 정답보기를 생성해야 한다. 이 경우 어휘의 의미적 중의성 문제가 발생할 수 있다. 예를 들어 동사 '내리다'는 '(탈 것에서) 타고 내리다'나 '(높은 곳에서) 오르고 내리다'를 포함한 다양한 의미를 가진다. 만일 유의어-반의어 사전을 그대로 이용한다면 "비가 내리다"의

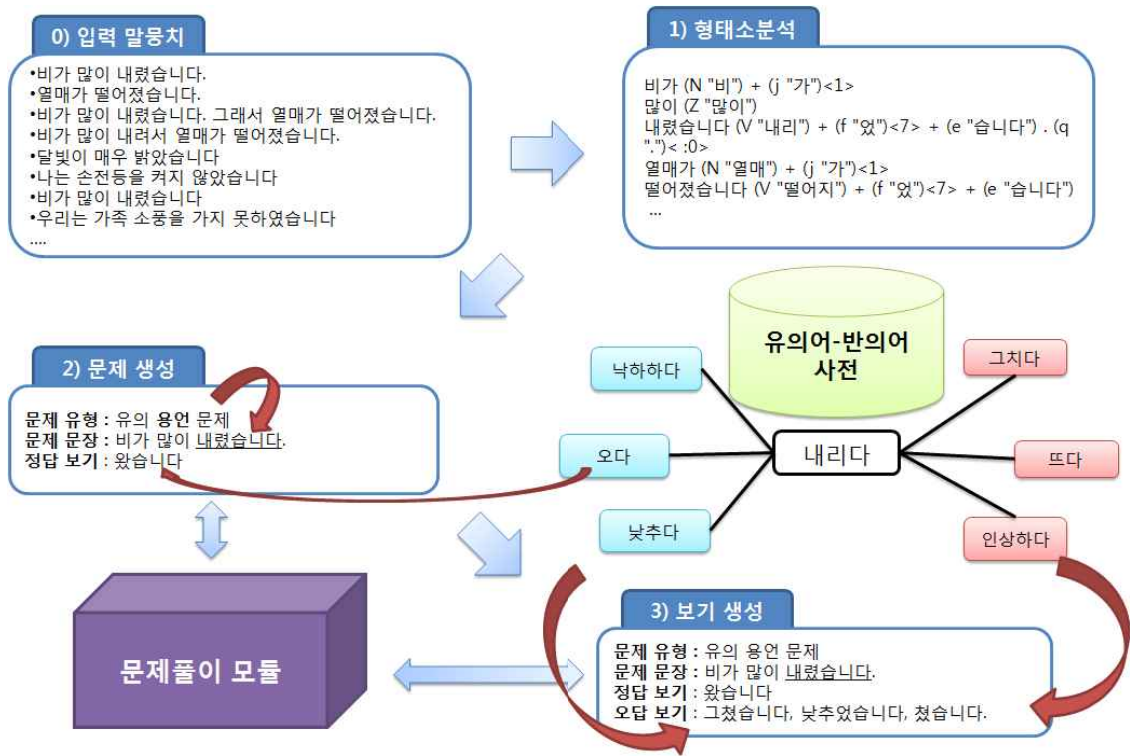


Fig. 2. Process of problem generation

'내리다'에 대해서 '낮추다'를 유의어 정답보기로 제시하는 경우가 발생할 수 있다. 본 논문에서는 이를 방지하기 위해 T8과 T10 유형에 대해서 문제 검증 기술을 적용해 적합한 정답 후보를 생성한다. Fig. 2에서는 “비가 많이 내렸습니다”에서 “내렸습니다”를 문제로 선택한 경우 유사 어휘로 “왔습니다”를 선택하는 예제를 보인다. '내리다'의 경우 유의어로 '오다', '낮추다', 반의어로 '뜨다', '그치다'를 가지는데, 이 어휘들 중 문장 “비가 많이 \_\_\_\_”의 공백을 자연스럽게 메꾸는 것은 '오다'가 되며, 이러한 자연스러운 어휘는 다음 절에서 설명하는 문제 검증으로 파악할 수 있다.

c) 사전과 말뭉치를 이용한 보기 생성

보기 생성에서는 생성된 문제와 정답보기를 바탕으로 오답보기를 생성한다. 오답보기는 문제유형에 따라 정답보기의 유의어와 반의어, 문맥에서의 공기 관계를 활용하여 생성한다. Fig. 3은 T2 적합 용언 문제에 대해서 문제와 정답 보기 생성을 통해 '놓치지'를 정답으로 선정한 뒤 나머지를 세 개의 오답보기를 선정하는 예를 보인다. 오답보기는 정답이 아닌 임의의 어휘로 쉽게 생성할 수 있으나, 만일 아래 예제의 경우 정답과 아무 상관이 없는 '필요하지'나 '깨지' 등을 보기로 제시하면 문제의 난이도가 낮아지는 문제점이 있다. 시스템에서는 '놓다'의 유의어인 '꿂다'와 반의어인 '당기다', '기회'와 자주 공기하는 '날다'를 오답보기로 생성하여 문제의 난이도를 보장한다.

보기 생성은 두 단계로 구성된다. 첫 단계에서는 정답에 대한 유의어-반의어 사전과 웹문서나 말뭉치에서의 공기 관

다음 ( )에 알맞은 것을 고르십시오.

보건소에서 무료 검사를 받을 수 있는 기회를 ( ) 마시기 바랍니다.

- ① 꿂지 <유의적 보기>      ② 날지 <체인과 연관된 표현>
- ③ 놓치지 <정답>          ④ 당기지 <반의적 보기>

Fig. 3. Example of generating an answer and distracters

계를 활용하여 세 개의 오답 후보를 추출한다. 추출된 어휘는 정답과 의미관계가 존재하므로 난이도를 보장할 수 있다. 만일 유의어나 반의어, 공기 관계로 적절한 후보를 얻을 수 없으면 임의의 추출을 통해 오답보기를 생성한다. 두 번째 단계에서는 선택된 어휘에 대해서 문제 검증을 시행한다. 예를 들어 Fig. 3의 예제에서 '잃지'가 오답보기로 선택되면 두 개의 정답이 가능한 상황이 발생한다. 검증 단계에서는 오답보기를 입력한 문맥이 말뭉치나 웹문서에서 자주 발생하면 이를 제외시켜 보기의 적합성을 보장한다. 제외된 오답보기가 발생하면 세 개의 오답보기가 생성될 때까지 보기 생성 과정을 반복한다.

3) 문맥의 빈도를 이용한 문제 검증

문제 검증에서는 보기후보를 문제의 문맥에 대입하여 만들어진 어구가 웹문서나 말뭉치에서 얼마나 많이 발생하는가의 여부로 정답이나 오답보기의 적합성을 검증한다. 이때 보기가 포함된 문장 전체의 발생 빈도를 사용하는 것이 가장 좋지만, 넓은 문맥을 사용하면 자료부족문제로 도리어

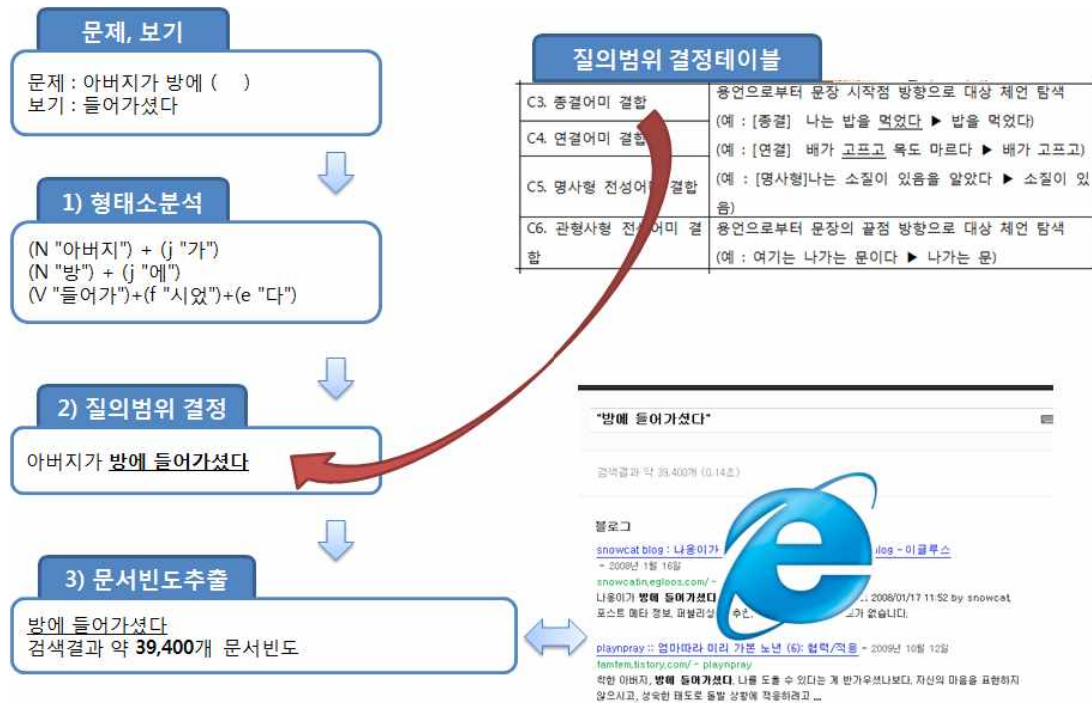


Fig. 4. Process of problem verification

정확률이 떨어질 수 있다. 본 논문에서는 정답의 품사와 형식 형태소와의 관계에 따라 각기 다른 범위의 두 어절을 질의어로 사용하여 자료부족문제를 해결하는 동시에 빠른 처리를 보장한다. 문제 검증은 형태소 분석된 문제에 대해서 Fig. 4와 같이 연관어절을 포함한 질의범위를 결정하고 이에 대한 문서 빈도에 기반하여 적합성을 검증한다.

a) 연관어절을 포함한 질의범위 결정

연관어절은 밑줄이나 괄호에 들어갈 어휘와 수식-피수식이나 지배-피지배 등의 연관 관계가 있는 어절을 의미한다. 연관어절은 정확한 구문 분석을 통해 얻을 수도 있지만, 초급 수준의 사지선다 문제의 경우 길이가 짧고 문장 구조가 단순하여 간단한 휴리스틱으로도 비교적 좋은 연관어절추출

이 가능하다.

본 논문에서는 Table 2와 같이 문제 어휘의 품사와 결합 형태소에 따라 각기 다른 방식으로 연관어절을 추출하여 검증에 사용한다. Table 2의 C1은 “그 학교의 학생들은 성실하다”나 Fig. 5의 “소비자들 사이에 큰 반응을 얻었다”와 같이 보기의 품사가 체언이면서 뒤에 조사가 결합하는 경우이다. 이 경우에는 해당 체언에 대한 서술 용언이나 피수식 체언을 우측 즉 문장 끝 방향에서 찾아 “학교의 학생들은”과 “반응을 얻었다”를 문맥으로 추출하는 것이 적합하다. C2의 용언화접미사가 결합된 체언은 문장 내에서는 용언으로 활용되는 것이므로 용언의 주어나 목적어 등을 찾기 위해 문장의 시작점 방향으로 진행하며 용언의 대상이 되는 체언을 탐색한다.

Table 2. Context extraction range for problem verification

품사	결합 형식형태소	질의범위 방향
체언	C1. 조사	체언으로부터 우측으로 행위, 서술 용언이나 피수식 체언 탐색 (예 : 나는 밥을 먹었다 ▶ 밥을 먹었다)
	C2. 용언화 접미사	체언으로부터 좌측으로 행위, 서술 용언 탐색 (예 : 근처로 이사했어요 ▶ 근처로 이사)
용언	C3. 종결어미	용언으로부터 좌측으로 대상 체언 탐색
	C4. 연결어미	(예 : [종결] 나는 밥을 먹었다 ▶ 밥을 먹었다) (예 : [연결] 배가 고프고 목도 마르다 ▶ 배가 고프고)
	C5. 명사형 전성어미	(예 : [명사형]나는 소질이 있음을 알았다 ▶ 소질이 있음)
	C6. 관형사형 전성어미	용언으로부터 우측으로 대상 체언 탐색 (예 : 여기는 나가는 문이다 ▶ 나가는 문)
조사	C7. 체언	체언의 조사 결합과 동일한 방향으로 질의범위 결정

보기 어휘가 용언인 경우에는 어미의 종류에 따라 연관어휘의 위치가 달라진다. 용언에 관형사형 전성어미 이외의 어미가 결합된 경우에는 보기로부터 좌측 즉 문장의 시작점 방향으로 용언의 대상 체인을 탐색한다(C3~C5). Fig 4의 문제 어휘인 '들어갔다'는 용언에 종결어미가 결합되었으므로 좌측 방향에서 체인을 찾아 "방에 들어갔다"를 질의로 추출한다. 이에 반해 '나가는'과 같이 관형사형 전성어미가 결합된 용언은 문장의 끝점 방향으로 용언의 대상 체인을 탐색한다(Table 2의 C6). 조사는 체인에 조사가 결합된 것과 동일한 방식으로 질의범위를 결정한다.



Fig. 6. Example of automatic word problem generation

---

**[문제 문장]** 소비자들 사이에서 큰 ( )을 얻었다.  
**[입력 보기]** : 반응  
**[형태소 분석결과 및 예상 지경범위]**  
 (소비자:N)(들:s) (사이:N)(에서:j) (크:V)(-:e) **(반응:N)(을:j)**  
**(얻:V)(었:f)(다:e)**  
**[최종 질의문]**  
 “반응을 얻었다”

---

Fig. 5. Example of deciding query range with related words

b) 문맥 빈도 추출

추출된 문맥의 실제 사용 빈도를 얻기 위해 말뭉치나 웹문서를 검색하여 문맥이 포함된 문서 빈도를 추출한다. 본 논문에서는 구글(<http://www.google.co.kr>)에 대한 웹문서 검색을 통하여 문맥의 빈도를 추출한다. 대부분의 검색 엔진에서는 “반응을 얻었다”와 같이 쌍따옴표를 사용하여 문서에서 질의어가 연결하여 나타난 빈도를 얻을 수 있다. 검색에서는 대량의 웹문서를 사용하므로, 자료부족을 해결하기 위해 “반응을 얻”과 같은 실질형태소를 사용하는 방법이 아닌, Fig. 5의 “반응을 얻었다”와 같은 원 문장 형태를 그대로 사용하는 방법으로 검증의 정확도를 높이고자 한다.

c) 보기 적합성 검증

정답보기의 적합성 검증에서는 웹 검색을 통해 얻어진 빈

도가 임계값을 넘으면 실생활에서 충분히 쓰이는 문장으로 보고 올바른 정답으로 간주한다. 초중급 수준의 한국어 두 어절은 대용량의 웹문서에 대한 검색에서 십만건 이상의 검색 결과를 가지지만 웹문서의 특성상 오타나 문법에 맞지 않는 표현들이 검색될 수 있어 본 논문에서는 경험적인 기준으로 1,000을 임계치로 사용한다.

오답보기의 적합성 검증을 위해서 오답보기가 대입된 문장이 실생활에서 충분히 쓰인다면 이는 정답일 가능성이 존재하므로 오답보기로 제시하지 않아야 한다. 본 논문에서는 (오답보기의 문서빈도)/(정답보기의 문서빈도)가 임계치를 넘지 않는 경우만 오답보기로 출제하여, 보기에서 두 개 이상의 정답이 발생하지 않도록 처리하며 본 논문에서는 0.15를 임계치로 사용한다.

Fig. 6은 제안된 방식으로 자동 생성된 문제를 실제로 풀어볼 수 있는 시스템의 화면을 보인다. 시스템은 관리자가 설정한 문서로부터 자동으로 문제를 생성하며, 사용자의 풀이 정보를 저장하여 내공부방에서 오답노트 등을 제공한다.

3.2 표준 발음 표기

표준 발음 표기에서는 한국어 읽기에 대한 학습을 지원하기 위해 국립국어원에서 제시하는 표준 발음법에 근거하여 임의의 문장에 대한 발음문자열을 생성한다. 표준 발음법상의 규칙은 인간이 이해하기 쉽도록 원칙, 음운체계, 변동의 순서로 체계화되어 있으며, 각 규칙들은 상호 영향을 미

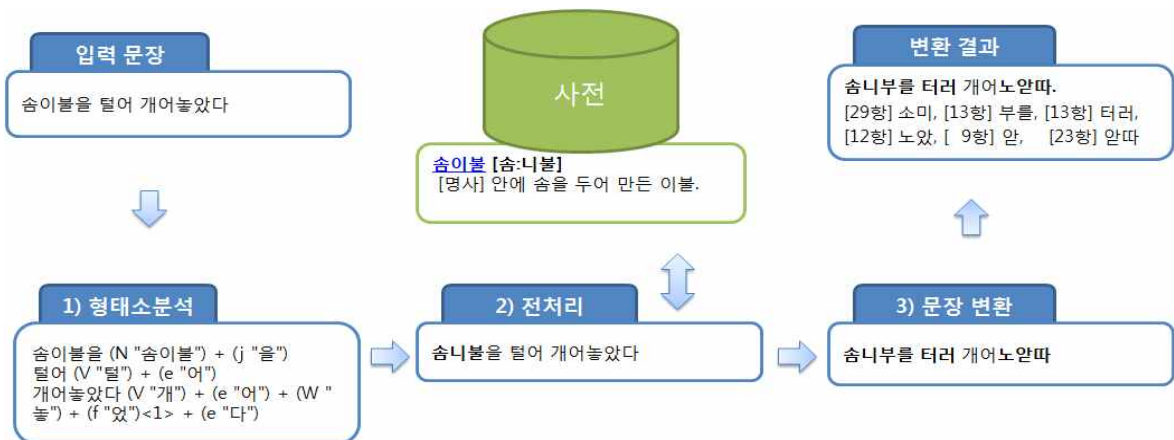


Fig. 7. Process of generating standard pronunciation marker

친다. 따라서 규정을 순서대로 적용하는 등의 방법으로는 올바른 발음 표기를 얻을 수 없다. 본 논문에서는 자동화된 시스템에서 표준 발음을 제시하기 위해서 규칙을 분석하여 그 순서를 재설정하고 Fig. 7과 같은 절차로 표준 발음 표기를 얻는다.

발음 규칙의 적용 순서를 설정하기에 앞서 표준 발음법의 2조 5항, 4장 9항부터 7조 30항까지의 규칙을 Table 3과 같이 네 가지 유형으로 분류하였다. 규칙을 유형별로 분류함으로써 규칙에 필요한 정보와 순서를 파악할 수 있다. 단순 변환의 경우 ‘[4장 9항] 평파열음화’와 같은 음절, 어절단위의 변환이며 별도의 정보를 필요로 하지 않는다. 문장 의존 변환의 경우 ‘[6조 24항, 25항] 어간 받침 처리’나 ‘[6조 27항] 관형사형 어미처리’ 등에 해당하며 형태소와 같은 추가 정보를 필요로 한다. 불규칙 변환은 ‘[4조 16항] 한글 자모의 특별한 발음’과 같이 일반 규칙과 상반되는 변환이다. 예를 들어 한글 자모의 이름에 대해서는 일반적인 연음을 적용하지 않아 ‘디글’의 경우 ‘디글이’는 ‘디그시’로 발음된다. 특수 변환은 ‘[6장 26항] 한자어의 된소리 발음’과 같이 사전에 의존하여야 하는 예외적 발음이다. 불규칙변환과 특수변환의 경우 일반적인 규칙과는 다른 내용을 담고 있기 때문에 우선적으로 처리하여야 한다. 1조 1항 - 2조 4항, 3조 6항 - 4조 8항의 경우 발음문자열로 변환하는 과정과 무관한 항목이며, 5조 21항, 22항의 경우 인정하지 않는 자음동화와 예외적 추가허용에 대한 규칙이므로 적용 대상에서 제외하였다.

발음문자열로의 변환에서는 문법적 특성과 음절에 따른 규칙적 변환을 반영하여 의미있는 자동화를 구현했다. 표준 발음법의 규칙을 임의의 순서로 적용하여서는 발음 문장으로 올바르게 변환할 수 없으므로 본 논문에서는 규칙 간 상관관계와 역행적 규칙적용의 관계를 따져 Fig. 8과 같이 전체 규칙을 재구성하였다. 불규칙 변환과 특수 변환은 규칙의 특성상 전처리단계에서 미리 처리한다. 변환 단계에서는 다른 규칙과 연관성이 없을 경우 표준 발음 표기법의 순서

Table 3. Class of pronunciation changes

변환 유형	변환 내용
단순 변환	음절, 어절단위의 단순한 변환
문장 의존 변환	문법적 특성을 고려하는 변환 (5, 13, 15, 17, 24, 25, 27항)
불규칙 변환	일반 규칙과 상반되는 변환 (16, 20항, 26항 다만)
특수 변환	사전에 의존하여야 하는 변환 (26, 28, 29항)

**전처리 단계**

한자어, 합성어, 파생어 등의 우선처리 (4장 16항, 5장 20항, 6장 26항, 28항, 7장 29항)

**변환 단계**

- ① 2장 5항 [단모음화]
- ② 4장 12항, 12항 붙임 [격음화]
- ③ **6장 24항, 25항** [경음화]
- ④ 6장 27항 [경음화]
- ⑤ **5장 17항 붙임, 17항** [구개음화]  
4장 13항, 14항, 15항 [연음, 불파음화], + **일반적 연음**
- ⑥ 5장 19항, **18항(역행동화)**, 20항 [비음화, 유음화]
- ⑦ 7장 30항 [경음화]
- ⑧ **4장 9항, 10항, 11항** [평파열음화]
- ⑨ 6장 23항 [경음화]

Fig. 8. Detailed process of generating pronunciation marker

를 최대한 유지하되 발음 현상의 특성과 시스템 처리의 효율성을 고려하여 순서를 조정한다. 예를 들어 종성이 다음 글자의 초성과 작용하는 5장 17항, 4장 13항-15항(구개음화, 연음, 불파음화)와 같이 서로 다른 규칙 항에 해당하지만 함



Fig. 9. Our pronunciation marking system

게 묶어 처리하는 것이 편리한 항은 Fig. 8의 단계 ⑤와 같이 함께 묶어 처리한다. 표준발음법에는 한 어휘 내에서 발생하는 일반적인 연음 규칙이 존재하지 않는다. 예를 들어 ‘입원’이 ‘이뵤’, ‘놀이’가 ‘노리’로 발음되는 것은 실제로 자주 발생하는 현상이다. 그러나 표준발음법상에서는 이러한 규칙이 존재하지 않아 한 어휘 내에서 발생하는 일반적인 연음을 처리할 수 없으므로, 시스템에서는 이를 항목 ⑤에서 추가로 처리하는 등의 방법으로 표준발음표기를 생성한다.

Fig. 9는 시스템의 발음표기 화면 예를 보인다. 입력된 한글 문장에 대한 발음 문장과 함께 로마자표기법을 제시하여 외국인의 읽기 학습을 지원한다.

### 3.3 글쓰기 학습지원

영어나 중국어 등에서는 나타나지 않는 다양한 조사와 용언의 활용은 한국어를 학습하는 외국인들이 가장 어려워하는 부분이다[13]. 글쓰기 지원 시스템에서는 문맥에 맞는 용언이나 조사를 추천하여 글쓰기에서 발생할 수 있는 오류를 최소화한다. 또한 문장에 포함되어야 할 어휘들을 나열하면 해당 어휘들을 포함하여 자주 사용되는 문장을 추천하여 글쓰기를 돕는다. 조사-용언의 추천과 유사 문장의 제시에서는 말뭉치의 공기 관계를 활용하고, 실시간 추천을 위해 기분석 사전을 데이터베이스에 저장하여 활용한다. Fig. 10에서는 글쓰기 학습지원의 화면 예를 보인다.

#### 1) 적합 조사와 용언 추천

적합 조사와 용언의 추천에서는 말뭉치의 공기 관계를 사용한다. Fig. 10의 조사-용언 추천 부분에서는 추천 예를 보인다. 사용자가 “예쁜 꽃이”를 입력한 뒤 단어 추천 버튼을 클릭하면 ‘꽃이’ 뒤에 가장 자주 발생하는 용언인 ‘핀다’와 ‘피다’ 등을 빈도 순서로 나열하면서 각 어근의 사전적 의

미를 제시하여, 글쓰기 지원과 동시에 어휘 학습을 지원한다. 만일 사용자가 “예쁜 꽃”을 입력한 뒤 추천을 원하면 시스템에서는 체언 ‘꽃’ 뒤에서 자주 발생하는 조사를 용언과 유사한 방식으로 나열한다.

본 논문에서는 자료부족문제를 줄이기 위하여 조사 추천은 체언 바로 뒤에 나타나는 조사로, 용언 추천은 ‘체언+조사’ 뒤에 나타나는 용언으로 제한한다. 검색 빈도만으로 구현할 수 있는 문제 검증과는 다르게 조사와 용언의 추천에는 수많은 경우의 수가 발생하는데 이를 말뭉치나 웹문서 검색으로 처리하기에는 시간적 부담이 크다. 본 논문에서는 실시간 처리를 위해서 기분석 자료를 활용한다. 기분석 과정에서는 세종말뭉치를 분석하여 체언과 조사의 결합 확률과 체언과 조사, 용언의 결합 확률을 색인된 형태로 데이터베이스에 저장하여 실시간 처리를 지원한다.

#### 2) 유사 문장 추천

Fig. 10 아래쪽의 유사문장 목록에서는 문장 추천의 예를 보인다. 사용자가 “예쁜 꽃이 핀다” 또는 “예쁘다 꽃” 등을 입력하면, 입력된 문장의 체언과 용언을 추출하고 해당 체언과 용언이 포함된 말뭉치의 문장을 화면과 같이 나열하여 원하는 의미의 자연스러운 한국어 문장을 작문할 수 있도록 지원한다.

유사 문장 추천에서도 색인된 기분석 데이터를 사용하여 실시간 처리를 지원한다. 기분석 데이터에는 원 문장과 문장에 포함되는 체언과 용언을 저장한다. 예를 들어 원 문장 “학교 꽃밭에는 예쁜 꽃들이 피어 있습니다”는 (‘학교’, ‘꽃밭’, ‘예쁘’, ‘꽃’, ‘피’, ‘있’)의 색인에 연결되어 있어 문장을 빠르게 추출할 수 있다. 시스템 부하를 줄이기 위해 문장에서의 중요성이 상대적으로 낮은 수식어는 기분석 데이터에서 제외한다. 유사 문장 판별에서는 의미적 중요성이



Fig. 10. Example of recommendation of predicate and similar sentence



좀더 높은 체언에 높은 가중치를 부여하고 용언에는 이보다 낮은 가중치를 부여하여, 사용자가 입력한 어휘와 각 문장에 포함된 어휘의 교집합에 대해서 “문장점수 = 일치 체언 수 × 1.0 + 일치 용언 수 × 0.7”로 유사 문장의 점수를 계산한다. 사용자에게는 유사도 점수 상위 10개의 문장을 제시하며 동순위 문장의 경우 짧은 문장을 우선 제시한다.

#### 4. 실험 및 평가

본 논문에서 제안하는 어휘문제생성, 문제 검증, 발음 표기 변환에 대한 평가를 시행하였다. 적합성 판정이 어려운 글쓰기 학습지원은 평가에서 제외하였다. 시스템 구축에서는 형태소분석기[16]와 mdict의 43만 유의어·반의어 전자사전[17], 형태소 분석된 세종말뭉치를 사용하였다. mDict은 유의어와 반의어가 단순히 나열된 형태이다. 문제 검증에서는 검색 엔진으로 구글을 사용하였다.

자동 생성된 문제의 적합성 평가에서는 초중급의 학습시스템의 수준에 맞게 입력 말뭉치로 초등학교 3학년에서 6학년에 해당하는 국민교과서 약 1,000 문장을 사용하였으며 자동 생성된 임의 400개의 문제에 대하여 수동으로 평가하였다. 평가에서는 문제 문장 및 정답에 대한 적합성과 오답 보기 3개에 대한 적합성을 별도로 평가하였다. 문제의 적합성은 문제유형에 따라 문제 문장이 올바르게 생성되고 정답 보기까지 올바르게 지정되었을 때 적합한 문제로 판단하였으며 ‘적합’, ‘부적합’으로 엄격히 판단하였다. Fig. 11은 문제와 정답 적합성 판정의 예를 보인다. 두 번째 예제의 경우

에 write의 의미로 쓰인 ‘적은’에 대해서 잘못된 의미 little의 반의어 ‘많다’를 사용한 보기 “많은 것”이 발생빈도가 많아 정답으로 선택한 오류를 보인다.

보기의 적합성 평가에서는 연관 어휘가 활용되면서 문제의 난이도를 보장하면 적합한 것으로 판단하였으며 정답이 될 수 있는 보기가 두개 이상이거나 연관 어휘가 사용되지 않아 문제의 난이도가 떨어질 때 부적합으로 간주하였다. Fig. 12는 판정 예를 보인다. 그림에서는 이해를 돕기 위해 정답은 모두 보기 (4)로 제시한다. 적합한 오답이 생성된 비율이 더 높았으나, 문제 검증의 오류로 복수 정답이 나오거나 유의어 사전의 부적합으로 난이도가 낮은 오답이 제시되는 경우가 있었다.

Table 4는 적합도 평가 결과를 보인다. 적합 어휘 선정 유형의 경우 생성된 문제와 정답보기에서 높은 정확도를 보였으며 오류는 대부분 형태소분석기의 오류에 의한 것으로 분석되었다. 유의 용언, 반의 용언 유형의 경우에는 원 문장에서 정답보기를 새롭게 생성하기 위해 유의어·반의어 사전을 참조하는 과정에서 나타날 수 있는 의미적 중의성 문제를 문제 검증을 통해 상당 부분 해소하였음을 알 수 있다.

Table 4. Suitability of problem and distracters

	적합 체언	적합 용언	적합 조사	유의 용언	반의 용언	평균
문제 적합률	96.0%	97.7%	97.5%	83.3%	70.0%	88.9%
보기 적합률	96.0%	53.4%	55.0%	50.0%	70.0%	64.9%

문제유형	문제	제시 정답	평가
T10 반의용언	얼음같이 차가운 바람이 나뭇가지 사이로 스친다	뜨거운	적합
T10 반의용언	앞 쪽에 적은 것 중에서 두 가지를 골라 봅시다.	많은	부적합

Fig. 11. Example of evaluating automatic generated answers

문제유형	문제	오답보기	평가
T1 적합체언	깨끗하고 아름다운 학교를 만들기 위해서 ( )을 합니다.	(1) 휴식 (2) 총력 (3) 촬영 (4) 일	적합
T2 적합용언	꼭 하고 싶은 일을 분명한 목소리로 ( ) 봅시다.	(1) 행동하여 (2) 털어 (3) 기뻐해 (4) 말하여	적합
T3 적합조사	선배들의 이야기( ) 들었습니다.	(1) 에는 (2) 와 (3) 로만 (4) 도	적합
T1 적합체언	학교 앞에 있는 ( )에 환자들이 많이 옵니다.	(1) 도적 (2) 건축 (3) 재앙 (4) 병원	부적합 (난이도)
T3 적합조사	학교 사육장에는 귀여운 동물( ) 있습니다.	(1) 에게 (2)처럼 (3) 도 (4) 이	부적합 (복수정답)
T6 유의용언	오랜 역사를 <지닌> 우리 민족	(1) 잠자다 (2) 다잡다 (3) 두들기다 (4) 가지다	부적합 (난이도)

Fig. 12. Example of evaluating automatic generated distracters

보기 적합률은 적합 체언 문제의 경우 96%로 나타났으나 적합 용언, 유의 용언, 반의 용언의 경우 50%를 약간 상회하는 낮은 정확도를 나타냈다. 이는 전자사전 상에 용언의 유의어와 반의어 정보가 충분하지 못하여 사용된 무작위 추출 보기로 난이도가 떨어진 경우가 대부분으로 파악되어 잘 구축된 온톨로지 등의 사용이 필요한 것으로 나타났다. 조사의 경우 55%의 적합률을 보였는데, 구어체에서 많이 쓰이는 ‘도’, ‘만’ 과 같은 보조사들이 문제 검증에서 오답으로 처리되지 않은 경우와, ‘있다’ 등의 다빈도의 다의 용언에 의한 경우들로 분석되었다.

본 논문에서 제안한 연관어절추출에 의한 문제 검증의 정확도를 평가하기 위한 실험을 수행하였다. 문제 검증은 웹문서 검색을 통해 문제의 적합성을 판별하는 것이므로 연관어절에 따른 문맥의 범위와 신뢰도는 밀접하게 연관된다. 실험에서는 각기 다른 질의범위의 크기와 방향을 사용하여 결과를 평가했다. Table 5는 실험 결과를 보인다. 표에서 R1은 보기를 대입한 문장 전체를, R2는 품사에 상관없이 앞뒤에 연결된 형태소 한개를, R3는 품사별 탐색방향을 고정된 방법(예를 들어 명사는 뒤에서 처음 나타나는 용언)을, R4는 본 논문에서 제안하는 방법을 질의범위로 사용한 결과다.

문제 검증 기술을 역으로 적용하여 잘 만들어진 문제에 정답보기를 대입한 것을 정답으로 판정하고 오답보기를 대입한 것을 오답으로 판정하는 비율로 제안한 문제 검증 기술을 평가하였다. 본 논문에서는 이를 각각 문제풀이의 정답신뢰도, 오답신뢰도로 구분한다. 실험에서는 TOPIK 기출문제와 본 논문에서 제안한 문제생성기술을 통해 생성된 문제를 합쳐 총 4,986개의 문제를 사용하였다.

Table 5. Reliability of problem verification

유형	R1	R2	R3	R4
정답신뢰도	2.1%	73.4%	71.2%	76.0%
오답신뢰도	98.0%	65.3%	69.2%	82.1%

Table 5의 R1과 R2에서는 정답신뢰도는 질의범위를 좁게 할수록 높게 나타나지만 오답신뢰도는 질의범위를 넓게 할수록 높게 나타나는 정답신뢰도와 오답신뢰도의 반비례관계를 발견할 수 있다. 이는 질의범위가 지나치게 짧으면 필요한 정보를 질의에서 제공할 수 없고, 지나치게 길면 자료부족문제가 발생하여 좋은 결과를 얻을 수 없기 때문이다. 반비례관계의 정답신뢰도와 오답신뢰도의 문제를 극복하기 위해 본 논문에서는 Table 5의 R4에 해당하는 연관어휘에 의한 의미적 질의범위를 제안하였다. 의미적 질의범위는 보기의 품사에 따라 품사와 밀접한 연관이 있는 실질형태소와 더불어 형식형태소를 포함한다. 실질형태소만을 질의에 포함시키는 R2와 R3의 경우에 검색에서의 자료부족문제 해소로 성능 향상을 기대할 수도 있으나, 실험에서는 형식형태소를 포함한 원형을 그대로 질의하는 방식이 문장 의미를 왜곡하지 않아 높은 신뢰도를 보였다.

발음표기의 성능평가를 위해 본 논문의 발음표기 시스템과 우리말 배움터의 표준 발음 변환기[8](이하 표준발음 변환기)를 비교분석하였다. 본 논문에서는 규칙의 자동화와 적용 순서에 초점을 두므로 한자어 및 합성어, 파생어에 해당하는 6장 26항, 28항, 7장 29항은 비교대상에서 제외하였다. 비교 대상 문장으로는 국립국어원에서 제공하는 세종말뭉치의 76만 원시문장 중 421문장 4389어절을 임의 추출하여 발음문자열로 변환하였으며, 표준발음법을 근거로 비교하였다.

Table 6은 발음 변환 결과에 대한 어절 단위 및 문장 단위에 대한 오류율을 나타낸다. 발음 문장으로 변환하는 과정에서 발생하는 오류는 어절 단위가 아닌 음운구 단위로 발생하므로 분석 단위를 음운구로 설정하는 것이 적합하다. 그러나 문장 내의 음운구를 분석하기 쉽지 않으며 변환기마다 다른 음운구의 단위를 가질 경우 비교 자체가 어렵다. 문장 내에 변환 오류가 발생할 경우 대체로 단일 음운구에서 오류가 발생하므로 어절 단위와 더불어 문장 단위로서 변환 오류도 함께 확인하였다. 예를 들어 ‘지어낸 얘기’를 변환하는 과정에서 이를 4장 13항(연음)에 해당한다고 잘못 판단하여 ‘지어내 내기’로 변환하였을 때 어절 단위로 오류를 판단하게 되면 두 어절에 걸쳐 오류가 발생한 것이 된다. 그러나 이것은 단일 오류에 대한 과잉 판단이다.

비교항목에 대해 6장 27항을 제외한 항목에 본 논문의 변환기가 상대적으로 높은 정확률을 보였다. 이에 대한 구체적인 예를 Table 7에 담았다. 표준발음 변환기는 대부분의 경우에 우수한 결과를 보이지만 실질, 형식형태소를 선정하는 과정에서 잦은 오류를 보인다. 이는 4조 13항, 4조 15항에 해당한다. 음운구의 설정으로 오히려 잘못된 결과를 낳기도 하였다. 예를 들어 6조 23항 ‘무척 좋은데’의 경우 경음화가 적용되지 않은 ‘무척 조은데’로 발음하는 것이 타당하다. 27항의 경우 본 논문의 변환기가 관형사형 어미 ‘-(으)르’을 올바르게 판단하지 못하여 상대적으로 낮은 정확도를 보였다.

Table 6. Evaluation of pronunciation marker

표준발음법 항목	어절 단위 오류 빈도		문장 단위 오류 빈도	
	표준 변환기	제안 변환기	표준 변환기	제안 변환기
4장 13항	25(0.5%)	0(0.0%)	17(2.0%)	0(0.0%)
4장 15항	42(0.9%)	0(0.0%)	25(2.9%)	0(0.0%)
5장 20항	17(0.4%)	0(0.0%)	17(2.0%)	0(0.0%)
6장 23항	34(0.8%)	17(0.4%)	34(4.0%)	17(2.0%)
6장 27항	17(0.3%)	33(0.8%)	17(2.0%)	33(3.9%)
총 오류빈도	135(3.0%)	50(1.1%)	110(13.0%)	50 (5.9%)

Table 7. Comparison of pronunciation markers

구분	원문장	표준변환기	제안변환기	설명
연음	들어오는 외풍	드러오느 너풍	드러오는 외풍	[4조 13] 홀받침이 모음으로 시작된 조사, 어미, 접미사와 결합 시 옮겨 발음
불과음화	눈 앞에서	눈 아페서	느 바페서	[4조 15] 받침 뒤의 모음 'ㅏ'로 시작되는 실질형태소가 연결 시 대표음으로 옮겨 발음 (예: 발 아래[바다래])
유음화	일을 놓지	이를 노치	이를 로치	[5조 20항] 'ㄴ'은 'ㄹ'의 앞이나 뒤에서 'ㄹ'로 발음 (예 : 바람 짚을 날 [바람자즐랄])
경음화	무척 좋은데	무척 쫘은데	무척 조은데	[6조 23항] 받침 'ㄱ, ㄷ, ㅂ' 뒤에 연결되는 'ㄱ, ㄷ, ㅂ, ㅅ, ㅈ'은 된소리로 발음 (예: 국밥[국뽵])
경음화	깔고 들어가는	깔고 드러가는	깔꼬 드러가는	[6조 27항] 관형사형 -(으)ㄹ 뒤에 연결되는 'ㄱ, ㄷ, ㅂ, ㅅ, ㅈ'은 된소리발음 (예: 만날 사람[만날 싸람])

### 5. 결 론

언어 학습을 위해 컴퓨터를 활용하는 교육 방식은 오래전 부터 연구되어왔다. 언어를 자연스럽게 구사하기 위해서는 많은 학습 시간을 필요한 만큼 언제 어디서나 지속할 수 있는 학습지원시스템은 매우 중요하다. 그러나 대부분의 한국어 학습시스템이 콘텐츠 위주로서 한정된 정보만을 제공하며 자동화 수준이 높지 않아, 언어의 동적인 측면을 빠르게 반영할 수 없고 한정된 내용만을 제공하고 있다.

본 논문에서는 적은 비용으로 학습지원시스템을 구축하는 방법을 제안하였다. TOPIK의 문제유형에 기반하여 말뭉치와 사진을 이용하여 자동으로 문제를 생성하고 문제의 적합성을 보장하기 위한 웹문서 검색을 통한 문제 검증 방식을 제안하였다. 발음 변환을 위해 발음 규칙 적용 순서를 정의하고, 글쓰기 지원을 위해 말뭉치에 기반한 조사-용언 추천과 문장 추천 방식을 제안하였다. 실험에서는 제안한 시스템이 문제와 정답 생성에는 높은 적합률을 보였으나, 용언과 조사에 대한 보기 생성에서는 유의어-반의어 사진의 내용 부족 등으로 낮은 적합률을 보였다. 이는 손쉽게 획득할 수 있는 지식원에서 자동 생성된 점을 감안하면 높은 수준으로 볼 수 있다.

실용적인 CALL을 구축하기 위해서는 자동 생성된 시스템에 의존할 수는 없다. 한국어에 대한 잘 구축된 온톨로지나 구문분석기 등을 습득하기 쉽지 않으나 이를 도입하여 제안된 시스템의 성능을 향상시키면, 비교적 적은 전문가의 투입으로도 실용적인 한국어 학습지원시스템을 구축할 수 있을 것으로 기대한다. 추후 연구로는 본 논문에서 다루고 있지 못하는 유형의 문제와 학습 요소에 대한 연구를 진행 중에 있다.

### 참 고 문 헌

[1] 김광해, “어휘력과 어휘력의 평가”, 선청어문 25권, pp.1-29, 1997.  
 [2] Scott, “An empirical study on explicit and implicit teaching

strategies in French”. The Modern Language Journal, pp. 14-22, 1989.

[3] Li, Lin Lin, “Analysis of Training Contents on Korean Educational Web Sites : For Chinese Learners of Korean Language”, MS thesis, Sangmyung University, 2010.  
 [4] Bae Si-yeong, “A system for learning English words using relations between words”, MS thesis, Hankuk University of Foreign Studies, 2010.  
 [5] Seong-Won Park and Chi-Hyung Lee, “Developing Educational Serious Game for English Learning”, Journal of Korean Society For Computer Game, Volumn 24, Issue 3, pp. 167-173, 2011.  
 [6] Sungjin Lee, Jonghoon Lee, Hyungjong Noh, Kyusong Lee, Gary Geunbae Lee, “Grammatical error simulation for computer-assisted language learning”, Knowledge-Based Systems, Vol.24, Issue 6, 2011.  
 [7] Chien-Liang Liu, Chia-Hoang Lee, Ssu-Han Yu and Chih-Wei Chen, “Computer assisted writing system”, Expert Systems with Applications, Vol.38, Issue 1, 2011.  
 [8] 우리말 배움터, <http://urimal.cs.pusan.ac.kr>  
 [9] Su-Il Choe, Ji-Hui Im, Ho-Seop Choe and Cheol-Young Ock, “Automatic Question Generation for Korean Word Learning System”, Proceedings of Spring Conference of Korean Cognitive Science, pp.9-14, 2006.  
 [10] 이규범, “국어 규범 교육을 위한 ‘표준 발음법’ 연구”, 석사학위 논문, 고려대학교, 2011.  
 [11] 국립국어원 표준발음법, [http://www.korean.go.kr/09\\_new/dic/rule/rule02.jsp](http://www.korean.go.kr/09_new/dic/rule/rule02.jsp)  
 [12] KANG Mi-Young and YOON Ae-Sun, “La génération de l’Alphabet Phonéique International pour l’apprentissage du coréen comme langue étrangée”, REVUE D’ETUDES FRANCAISES, Vol.51, pp.1-22, 2005.  
 [13] Yu Mi Kim, “A study of Error Analysis of Korean learners by using “Learner Corpus””, Teaching Korean as a Foreign Language, Vol.27, pp.141-168, 2002.  
 [14] Nam HyeonSook, Kwon HyunJu, Kim SuNam, Yu SungHun,

Sim MinKyu and Kwon Hyun-Chul, "Web-Based Intelligent Learning System for Korean Language", Proceedings of the 11th Conference on Hangeul and Korean Language Information Processing, pp.413-417, 1999.

- [15] Lee, Jeong Ryoung, "Design and implementation of a web-based system for creative writing instruction", MS thesis, Korea National University of Education, 2005.
- [16] Seung-Shik Kang, HAM, "Korean Morphological Analysis and Hangeul Analysis Module", <http://nlp.kookmin.ac.kr>.
- [17] mdict synonym and antonym dictionary, <http://www.octopus-studio.com>, Octopus Studio.



### 김 병 만

e-mail : [bmkim@kumoh.ac.kr](mailto:bmkim@kumoh.ac.kr)  
 1987년 서울대학교 컴퓨터공학과(학사)  
 1989년 한국과학기술원 전산학과 (공학석사)  
 1992년 한국과학기술원 전산학과 (공학박사)

1992년~현 재 금오공과대학교 컴퓨터소프트웨어공학과 교수  
 1998년~1999년 미국 UC, Irvine 대학 방문교수  
 2005년~2006년 미국 콜로라도 주립대학 방문교수  
 관심분야: 인공지능, 정보검색, 정보보안



### 박 기 태

e-mail : [newpark210@naver.com](mailto:newpark210@naver.com)  
 2006년~현 재 금오공과대학교 컴퓨터 소프트웨어공학과 학부과정  
 관심분야: 자연언어처리, 언어공학, 정보 검색



### 이 현 아

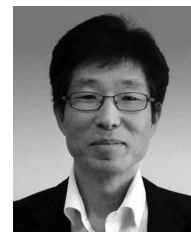
e-mail : [halee@kumoh.ac.kr](mailto:halee@kumoh.ac.kr)  
 1996년 연세대학교 컴퓨터과학과(학사)  
 1998년 한국과학기술원 전산학과 (공학석사)  
 2004년 한국과학기술원 전산학과 (공학박사)

2000년~2004년 (주)다음소프트 언어연구팀장  
 2004년~현 재 금오공과대학교 컴퓨터소프트웨어공학과 부교수  
 관심분야: 자연언어처리, 정보검색, 지식공학



### 이 태 훈

e-mail : [aliceland4@naver.com](mailto:aliceland4@naver.com)  
 2006년~현 재 금오공과대학교 컴퓨터 소프트웨어공학과 학부과정  
 관심분야: 정보검색, 언어공학 등



### 신 윤 식

e-mail : [ysshin@kumoh.ac.kr](mailto:ysshin@kumoh.ac.kr)  
 1982년 경북대학교(학사)  
 1985년 한국과학기술원 전산학과 (공학석사)  
 1988년~현 재 금오공과대학교 컴퓨터 소프트웨어공학과 교수

관심분야: 소프트웨어공학, 데이터베이스



### 황 소 현

e-mail : [ghkdthgus9@naver.com](mailto:ghkdthgus9@naver.com)  
 2009년~현 재 금오공과대학교 컴퓨터 소프트웨어공학과 학부과정  
 관심분야: 정보검색, 언어공학 등