

웹기반 언어 학습시스템을 위한 한국어 철자/문법 검사기의 성능 향상*

Improving a Korean Spell/Grammar Checker for the Web-Based Language Learning System

남 현 숙* 김 광 영* 권 혁 철**
(Hyun-Suk Nam) (Kwang-Young Kim) (Hyuk-Chul Kwon)

요약 이 논문의 목적은 한국어 철자/문법 검사기를 교육적으로 활용한 웹 기반 국어 작문 학습 시스템의 구현이다. 웹 기반 학습시스템 '우리말 배움터'의 학습효과를 최대화하려면 한국어 철자/문법 검사기의 성능을 꾸준히 향상해야 한다. 오늘날 자연어처리 시스템의 성능은 의미처리를 얼마나 정확하게 수행하는가에 달려있다. 한국어 철자/문법 검사기에서 의미처리와 관련이 있는 부분은 철자 검사기에서 접사나 꼬리말과 파생하는 단어와 복합명사를 교정하는 처리기와 의미·문체 오류를 교정하는 문법 검사기이다. 본 시스템에서는 의미처리를 위하여 의존문법에 기반하여 부분문장분석과 언어관계정보를 이용한다. 여기에 더 세부적인 규칙을 추가하기 위해 단어를 개념적으로 분류하고 문장의 핵심요소인 동사를 하위범주화한 결과를 적용한다.

의미처리 기능을 강화한 철자/문법 검사기를 온라인으로 운영함으로써 웹에 기반한 한국어 학습시스템과 통합된 환경에서 능동적이고 지능적인 학습 모형을 구현한다.

이 논문에서 다루는 의미처리의 대상은 주로 구문 단위이기 때문에 여러 개의 절이 모여 하나의 문장이 된 복문이나 중문은 다루지 못하고 있다. 또한 일률적인 체계 속에서 단어를 의미적으로 분류하는 데에도 많은 한계가 있다. 한편, 이러한 자연어처리시스템을 웹 기반 학습시스템에 연결하여 효율적인 학습효과를 거두려면 학습내용 구성이나 인터페이스 설계 면에서도 고려해야 할 중요한 문제가 많다. 결론에서는 아직 완전하게 해결하지 못한 문제에 대해 고찰한다.

키워드 한국어 철자/문법 검사기, 웹기반 학습시스템, 의미처리, 의존문법, 부분문장분석, 언어관계정보, 하위범주화

Abstract The goal of this paper is the pedagogical application of a Korean Spell/Grammar Checker to the web-based language learning system for Korean writing. To maximize the efficient instruction of our learning system 'Urimal Baeumteo', we have to improve our Korean Spell/Grammar Checker.

Today the NLP system's performance depends on its semantic processing capability. In our Korean Spell/Grammar Checker, the tasks accomplished in the semantic level are: the detection and correction of misused derived and compound nouns in a Korean spell-checking device, and the detection and correction of syntactic and semantic errors in a Korean grammar-checking device.

We describe a common approach to the partial parsing using collocation rules based on the dependency grammar. To provide more detailed semantic rules, we classified nouns according to their concepts, and subcategorized verbs referring to their syntactic and semantic features. Improving a Korean Spell/Grammar Checker makes our learning system active and intelligent in a web-based environment.

We acknowledge the flaws in our system: the classification of nouns based on their meanings and concepts is a time consuming task, the analytic unit of this study is principally limited to the phrases in a sentence, therefore, the accurate parsing of embedded sentences remains a difficult problem to solve.

Concerning the web-based language learning system, it is critically important to consider its interface design and structure of its contents.

1. 서론

철자 교정이 단순히 오류어를 찾아내는 것만을 목표로 하는 데에 반해 부산대학교에서 개발한 철자검사기의 교정시스템은 찾아낸 오류어에 대해 원래 사용자가 의도했던 단어를 찾아내어 대처하는 것을 목표로 하였다. 따라서 이 시스템은 대개 대화식으로 구성되어 제시한 대처어 중에서 가장 적합한 단어를 사용자가 직접 선택할 수 있도록 설계하였다. 한국어 철자/문법 검사기의 이러한 설계 취지는 웹 기반 국어 작문 학습시스템과 연결하여 자연어처리 시스템을 교육적으로 활용하고자 하는 이 논문의 연구목적으로 귀결한다.

이 논문에서는 철자/문법 검사기의 본래 취지에서 한 발 더 나아가 사용자가 스스로 대처어를 선택한 후 자신이 범한 오류와 관련이 있는 내용이 있는 페이지로 찾아가서 스스로 학습할 수 있도록 웹 환경에서 '우리말 배움터' 학습시스템과 통합한 한국어 철자/문법 검사기를 소개한다. '우리말 배움터' 학습시스템은 우리가 글을 쓸 때 필요한 우리 글에 대한 지식을 전달하려고 우리 나라 사람이 자주 틀리는 오류를 중심으로 예제에 기반한 글쓰기 학습을 진행해 나가도록 구성하였다. 학습시스템과 한국어 철자/문법 검사기는 철자/문법 검사기의 도움말 사전을 통해 내부적으로 연결된다. 도움말에서 분류되어 있는 오류 유형 즉, 우리가 자주 범하는 우리말 오류 유형에 기반하여 학습시스템의 내용을 제공한다. 학습시스템의 학습효과는 한국어 철자/문법 검사기의 성능과 밀접

한 관계를 맺는다.

이 논문의 본문에서는 주로 한국어 철자/문법 검사기에서 특히 의미처리를 어떻게 하는지에 대해 전개한다. 우리말의 특성을 고려하여 본 시스템에서 채택한 방법에 대해 설명하는데 이 논문에서는 다양한 의미처리 방법 중에서도 의존관계에 있는 서술어와 논항, 수식어와 피수식어 등 단어와 단어 사이의 관계를 주요 분석대상으로 삼는다.

한국어 철자/문법 검사기에서 의미처리와 관련이 있는 부분은 첫째, 어휘 중의성 제거 규칙과 부분 문장 분석을 이용한 의미·문체 검사기, 둘째, 언어 관계를 이용하여 접미사와 꼬리말을 처리하는 철자 검사기와 셋째, 의미정보에 따라 복합명사의 결합제약 규칙베이스를 구축하는 복합명사 처리기 등 세 가지이다. 이 각 처리기에서 수행하는 통사와 의미단계의 어절분석과정과 처리결과를 기술한다.

본 시스템에서 의미처리를 위해 기본적으로 의존문법에 바탕을 둔 부분문장분석과 언어관계정보, 의미계층구조를 부분적으로 활용한 어휘정보 등을 적용한다. 이를 위해 동사를 하위범주화하고 명사를 개념적으로 분류한다. 이렇게 의미처리 기능을 강화한 한국어 철자/문법 검사기를 웹 기반 국어 작문 학습시스템에 활용하여 더 나은 교정 결과와 더 효율적인 한국어 학습효과를 거두고자 한다.

2. 자연어처리시스템과 학습시스템의 통합환경

2.1 전체 구조

웹에 기반한 학습시스템은 인지주의 학습 이론에 바탕을 둔 학습 환경을 제공(delivery)할 수 있는 강력한 수단으로 등장했다. 인지주의자들은 어떠한 학습 환경에서도, 학습자가 구성하는 즉, 학습하는 지식의 유형에 따라 암기, 직접 교수, 연역법, 반복연습(drill and practice), 귀납법 등을 포함하는 다양한 학습 전략들이 요구된다고 주장한다(1, 12).

* 이 논문은 1999년 한국학술진흥재단의 연구비에 의하여 연구되었음. (KRF-99-E011)

* 부산대학교 자연대학 전자계산학과 인공지능연구소,

hsnam, kykim@solge.cs.pusan.ac.kr

** 공과대학 컴퓨터·정보통신학부

hckwon@hyowon.pusan.ac.kr

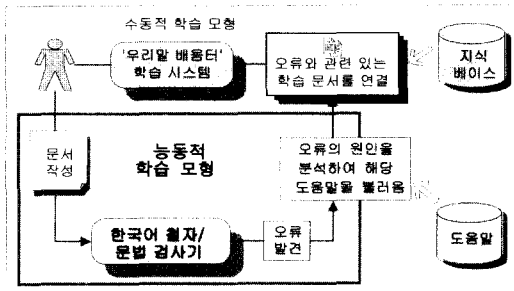
연구세부분야 : 한국어 정보 처리, 웹 기반 한국어 학습시스템

주소 : (우편번호 609-735)

부산시 광역시 금정구 장전동 산 30 (부산대학교)

자연대학 전자계산학과 인공지능연구소

전화번호 : 051-516-6764, 051-510-2875



(그림 1) 학습시스템과 한국어 철자/문법 검사기

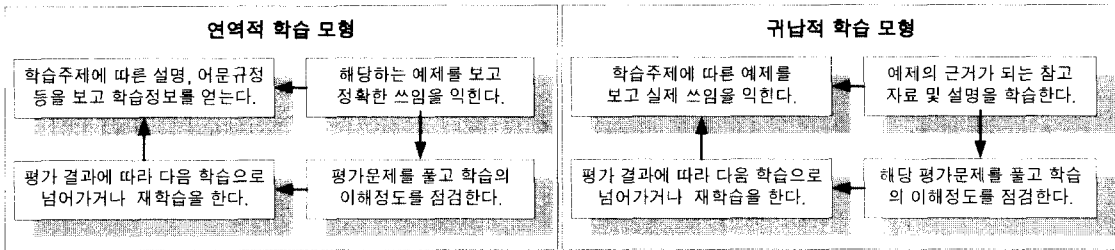
'우리말 배움터'는 학습자가 원하는 학습 정보를 찾아서 학습하는 수동적 학습 모형과 우리말 철자/문법 검사기를 통해 학습자가 스스로 실력을 평가해 보고 문제를 해결하는 능동적 학습 모형을 동시에 만족하도록 설계한 시스템이다.

즉시 확인해 보고 필요한 학습문서를 찾아가서 학습하도록 도와주는 일종의 자율 평가 시스템¹⁾으로 활용한다.

2.2 기존 학습시스템과 비교

본 시스템은 수동적 학습 모형과 능동적 학습 모형의 이중적인 구조를 띤다. 한국어 철자/문법 검사기를 연결하기 이전의 시스템은 (그림 1)에서 수동적 학습 모형을 구현한 것이며 현재 시스템은 능동적 학습 모형을 추가한 것이다. 수동적 학습 모형에서는 사용자가 단순하게 학습 내용을 선택하여 학습 정보를 익히는 데에 반해 능동적 학습 모형에서는 사용자가 철자/문법 검사기에서 글쓰기를 해 보고, 여기서 분석 결과에 따른 도움말을 보거나 더 상세한 지식을 알고자 한다면 학습시스템을 통해 심화 학습을 할 수 있다.

언어학습과 관련이 있는 자료나 사이트를 참고해



(그림 2) 연역적 학습 모형과 귀납적 학습 모형

흔히 시스템을 개발하는 사람들이 기술 경쟁에만 몰두한 나머지 대부분 시스템 성능의 우수성에만 신경을 쓰고 사용자 인터페이스에 관심을 쏟지 않는 잘못을 범하기 쉽다(9). 학습사이트에서는 단순한 정보를 제공하는 웹사이트와는 달리 정보를 제공함에 있어서 효과적이며 정확하게 전달하기 위해서 지식을 가공하고 적절하게 난이도를 조절해야 한다. 사용자가 가장 쓰기 편하고 학습할 내용을 쉽게 찾아서 익힐 수 있도록 설계한 시스템이 경쟁력이 있다. 과거 처럼 교사가 잘 정리한 지식을 학습자에게 주입시키는 방법은 설득력이 거의 없다. 따라서 학습자의 능력이나 태도에 따라 차이는 있을 것이나, 학습자 스스로 학습을 계획하고, 추진하며, 스스로 개발하도록 교육할 필요가 있다(2).

한국어 철자/문법 검사기는 학습자가 어느 단계에서든지 접속하여 문장을 써 보고 그에 따른 결과를

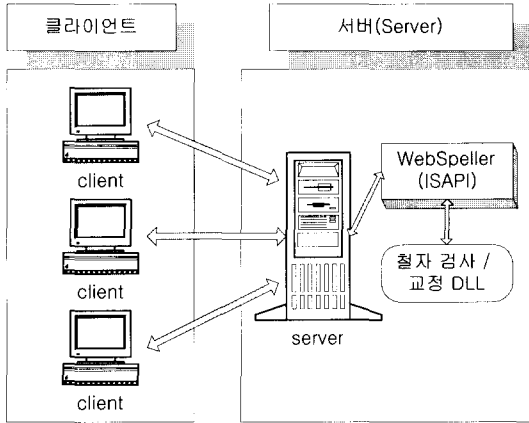
보았을 때 가장 자주 쓰는 논리 전개 방법은 '연역적 방법'과 '귀납적 방법'이다. 기초 지식이 없어도 입문서처럼 볼 수 있는 자료는 설명이나 풀이를 먼저 제공하고 예제를 보이는 '연역적' 전개 방법을 택하고, 기본 지식을 어느 정도 갖춘 사람이나 전문가를 위한 학습 자료는 '귀납적 전개 방법'을 택한다. 이와 같은 분석에 착안하여, 기존 학습시스템도 '연역적 학습 모형'과 '귀납적 학습 모형'을 적용하여 학습자가 수준에 맞게 선택하여 학습을 할 수 있도록 설계하였다.

이와 같은 두 가지 학습 모형 중에서도 현재 시스템은 귀납적 학습 모형을 더욱 발전시킨 형태로 나아간다. 즉, 사용자는 제시된 예제를 보는 것이 아니라

1) 자율평가시스템은 강의 중 어떤 때라도 학생들이 자체적으로 유연성 있게 평가할 수 있도록 즉각적이고 자동적이며 일관성 있는 피드백이 이루어지도록 하는 것이다(3).

스스로가 쓴 예제를 통해 근거 자료를 학습하게 되었다. 한국어 철자/문법 검사기를 학습 시스템과 통합은 기존 학습 시스템이 사용자에게 학습정보를 제공하는 데에 비중을 두는 수동성에서 탈피함으로써 사용자의 능동적인 학습활동을 꾀한다.

2.3 시스템 통합에 사용한 기술



(그림 3) 웹 기반 한국어 철자/문법 검사기

웹 기반 한국어 철자/문법 검사기는 ISAPI를 이용하여 클라이언트에서 서버로 동일한 요청에 대해 하나의 DLL만 메모리에 로드하고 새로운 요청이 있을 때마다 윈도우용 프로그램과 웹 프로그램 사이를 연동할 수 있도록 매개 역할을 하는 ECB (Extension Control Block) 구조체만 할당하여 사용한다. 다시 말해, 웹 스펠러 (ISAPI Extension)에서 철자 검사/교정기 DLL에서 문서 검사를 요청하면, 철자 검사/교정 시스템에서 검사 및 교정을 하여 그 결과를 다시 웹 스펠러 (ISAPI)에 전달하여 결과를 사용자에게 보여주도록 처리하였다.

2.4 두 시스템간의 학습내용 통합

한국어 철자/문법 검사기의 도움말은 우리나라 사람이 자주 틀리는 오류를 언어학적 원리(형태, 문법, 의미, 문체)와 언어 정책적 원리(한글 맞춤법, 표준어 규정, 외래어 표기법)에 따라 분류하였다. 학습시스템의 내용은 언어 분석 단위에 따라 맞춤법, 어휘, 문장, 문단 학습의 체계로 짜여져 있다. 오류 유형에 따른 도움말과 학습시스템의 내용은 논리적인 연결구조를 취한다(4).

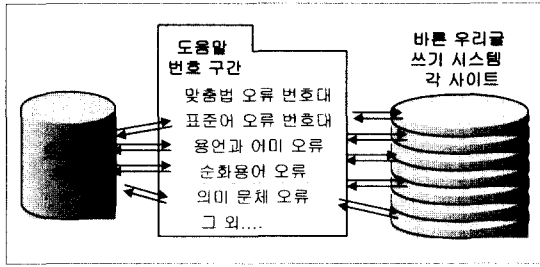
도움말을 통해 사용자는 자신이 범한 오류 유형에 따라 해당 학습사이트에 가서 학습할 수 있도록 각 도움말 번호가 구간별로 학습사이트에 연결되어 있다.

이 학습시스템의 메뉴는 우리말 학습의 기본인 맞춤법 학습(맞춤법 다지기)에서 어휘 학습(어휘 다듬기), 문장 학습(문장 다듬기), 문단 학습(문단 다듬기)으로 나아가는 구조이다. 무한한 웹 공간에서 정보를 조직적으로 구성하여 체계적으로 전달하기 위해서는 잘 계획된 라벨링 시스템²⁾ (planned labeling system)을 취해 학습자가 제목만 보고도 학습문서와 내용을 미리 짐작하여 가상 공간 내에서 길을 잃지 않고 학습을 마치도록 '학습의 맥락', '학습 과제의 방향성'을 부여한다. 한국어 철자/문법 검사기를 국어 작문 학습시스템에 연결한 설계 취지는 '실제적 설계'(6) - 학습자가 직접 글쓰기와 같은 실제적인 활동을 하게 하여, 현실적인 시나리오 안에서 상황적 연습과 피드백이 이루어지도록 하는 - 와 일맥상통한다.

(표 1) 철자/문법 검사기 도움말 사전과 학습내용 분류

처리 기	한국어 철자/문법 검사기 도움말	'우리말 배움터' 학습시스템 메뉴
철자 검사기	맞춤법 오류, 띄어쓰기 오류	맞춤법 다지기: 받음과 표기, 단어의 짜임새, 띄어쓰기
	유사 의미 오류, 유사 발음 오류, 표준어 오류, 외래어 표기, 순화 용어 오류	어휘 다듬기: 발음이 비슷한 말, 뜻 구별해야 할 말 자주 틀리는 표준어, 외래어와 우리말 순화
문법 검사기	문법 관련 도움말: 용언과 어미, 체언과 조사 관련, 서술어 문장성분 관계 도움말	문장 다듬기: 문장성분 (주어, 목적어, 서술어) 호응관계, 피동형과 사동형 문장 맞춤법 다지기: 단어의 짜임새 (체언과 조사, 용언과 어미 결합)
	좋은 문장과 나쁜 문장 도움말, 의미와 문체 관련 오류, 문장 구조 및 길이 관련 오류	문장 다듬기: 번역투 표현, 문장성분 호응관계, 문단 다듬기: 접속어의 바른 사용, 지시어의 바른 사용,

2) (...) 사용자들은 대체로 체계적이고 일관성 있는 라벨링 시스템이 어떤 분야를 나타낼 수 있는 한도에서는 최대한 범위 까지 보여줄 수 있다고 생각한다. (...) 한눈에 척 보면 사용자는 이 시스템이 어떻게 작동하는지를 이해하게 된다(5).



(그림 4) 도움말사전과 학습사이트 연결

3. 한국어 철자/문법 검사기에서 의미 처리

사실상 자연언어처리 시스템의 성능은 단어의 의미와 어휘 범주나 문장의 통사적 구조와 의미의 중의성을 얼마나 잘 처리하는가에 따라 결정된다(7).

한국어는 자유로운 어순과 불완전한 문장을 허용한다는 구문적 특징 때문에 한국어 구문분석에는 단일화(unification)에 기반한 구문분석과 의존문법에 기반한 구문분석이 주목을 받아왔다(8, 9). 의미론을 문법적인 관계나 문법적인 의존관계의 집합으로 보는 접근 방법이 있다. 하나의 문법으로 생성되는 의존관계의 집합은 서로 매치되는 패턴을 발견하고 그 결과의 총합을 하나의 논리적인 형태로 만들어서 수집하는 것으로 해석될 수 있다.

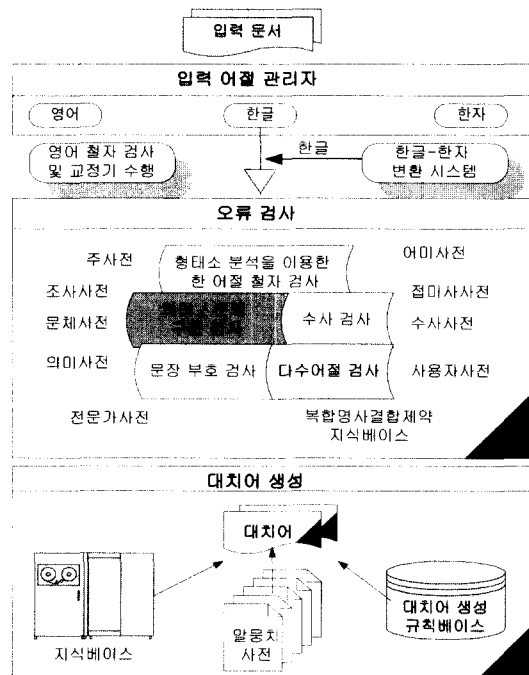
부산대학교에서는 90년대 초부터 한글과컴퓨터사와 한국어 맞춤법검사를 개발하여 상품화하기 시작하였는데 초기에는 주로 한국어 형태소분석과 태깅 기술에 대한 연구가 처리속도와 분석정확도 향상이라는 측면에 초점을 맞추어 수행되었다. 최근에는 한국어 문법 검사기의 개발에 활용한 의존관계에 의한 한국어 의미 처리 기법은 특정 언어에 의존한 처리 방법이라는 한계점에도 불구하고 의존관계에 기반한 규칙을 추가함으로써 시스템 성능을 지속적으로 개선할 수 있다(9, 10).

3.1 한국어 철자/문법 검사기의 전체 구성

본 연구실에서 개발한 한국어 철자/문법 검사기는 크게 어절 버퍼 관리자, 구 단위 철자 검사기, 구 단위 철자 교정기, 고빈도 말뭉치 사전을 이용한 대치어 검증부, 구 단위 연어 규칙 처리부, 문장부호 처리부, 문체 검사부로 구성되어 있다.

먼저, 어절 버퍼 관리자는 입력 문서에서 어절 단위로 입력을 받아 각 처리부에서 참조하기 쉽도록 토

큰(token) 단위로 분리한다. 이렇게 분리된 토큰은 환형 큐(circular queue)로 구성된 다수 어절 버퍼에 저장된다. 그런 다음 한 어절 버퍼링이 끝나면 먼저 한 어절단위 철자 검사/교정기가 오류 여부를 검사하고 철자 검사 과정에서 연어 정보 검사가 필요한 어절로 인식이 되면 연어 규칙 처리부가 구동한다. 한 어절의 분석이 끝난 후 그 어절의 앞뒤나 어절 사이에 문장 부호가 존재하면 문장 부호 처리부가 구동한다. 철자 검사기와 철자 교정기에서 생성된 대치어는 한국어 대용량 말뭉치에서 추출한 고빈도 어절 사전을 이용하여 검증한다.



(그림 5) 한국어 철자/문법 검사기 구성

문체 검사부는 인용부호 검사가 끝난 후 문장단위로 재분석하면서 문체 오류 지식베이스를 검색한다. 문체 오류 가능성이 있는 단어가 발견되면 지식베이스에 따라 그 제약 조건을 검사하여 문체 오류 여부를 판단한다. 문체 오류는 대치 문장을 생성하는 것이 어려우므로 오류 문장을 반전하고 오류에 대한 도움말을 제시한다.

분석 단위가 구문이나 문장인 문법 검사기의 의미 처리 과정을 기술하기에 앞서 한 어절 내에서 의미처

리를 수행하는 철자 검사기에 대해 먼저 알아본다.

3.2 철자 검사기

3.2.1 접미사와 꼬리말 교정

철자 검사기에서 분석 대상은 주로 하나의 어절 단위이므로 형태론적 단계에서 다루는 내용이 대부분이며 구문분석에 기반한 한국어 처리의 기초가 되는 기술이다 [22]. 여기에서 어휘의미적인 정보를 비교적 많이 적용하는 범위는 접미사와 결합하는 파생어 영역이다.

〈표 2〉 철자 검사기에서 의미결합정보를 이용한 접미사와 접미명사 분류

접미사, 접미명사	의미 결합 정보
-류, -값	동물이나 식물 명사와 결합 가능
-깃, -꽂지, -동우리, -동지, -부리	동물(조류)와 결합 가능
-탕, -잡이, -치어, -회	동물(어류)와 결합 가능
-답다, -무리, -떼	주로 사람명사와 결합, 동물명사 결합 가능
-즙, -씨, -가시, -술, -울, -뿌리, -잎, -꽃, -줄기	식물 명사 뒤에만 붙는 접미사와 접미명사
-어(언어), -인(사람)	나라 이름 명사와 결합 (예, 한국인, 한국어)
-계, -회	'-학'으로 끝나는 학문을 나타내는 명사
-변	강 이름을 나타내는 명사와 결합
-팀	사람, 직급과 지역, 나라, 도, 시 이름을 나타내는 장소 명사

접미사가 다른 단어와 결합을 할 때 의미적 분류 정보에 따라 다르게 붙는다. 철자 검사·교정기는 접사 또는 접사의 특성을 띠는 명사를 '접미·접두 명사'³⁾라는 어휘범주를 따로 설정하여 처리하고 '꼬리

말 사전'이라는 부분사전을 관리한다. 〈표 2〉는 접미사와 다른 단어를 의미적으로 결합하는 규칙의 예를 나타낸다.

접미사가 붙어 의미가 바뀌는 명사 중에는 지명을 가리키거나 직책을 가리키는 예가 가장 많다. 의미적 분류 정보에 따라 지명과 언어관계가 성립하는 접미사나 다른 명사를 검사한다. 또한 접미사가 어떻게 붙는가에 따라 그 단어의 전체 의미가 달라진다. 아래의 표는 접미사가 다른 명사와 어색하게 결합한 예를 보여준다.

이와 같은 접미사와 명사의 의미적인 결합 제약 때문에 〈표 2〉에서 보는 바와 같이 장소명사는 지역, 나라, 도, 시 등의 명사로, 사람명사는 직책이나 직업과 관련이 있는 명사로 세분화하였다.

〈표 3〉 접미사와 명사의 의미적 결합 오류의 예

접미사	맞는 어절	틀린 어절
-어	한국어, 일본어, 중국어	서울어, 부산어
-청	부산시청, 강원도청	미국청, 일본청, 가락동청
-인	한국인, 미국인, 일본인, 중국인	서울인, 부산인, 제주인
-회	어머니회, 노인회, 어린이회	학교회
-장	학교장, 단체장, 연구회장	기능공장

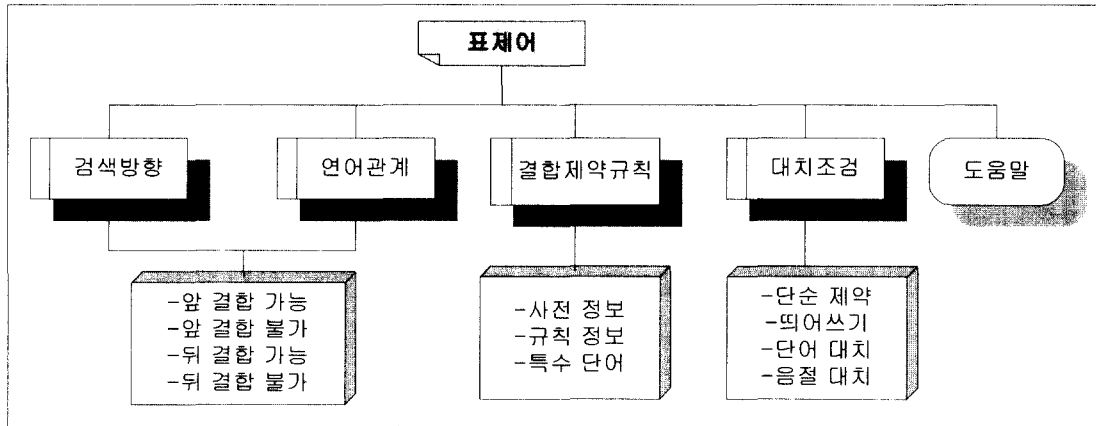
3.2.2. 복합 명사 교정

복합명사의 의미 처리는 하위범주화⁴⁾한 어휘정보를 수록한 전자사전을 이용하여 수행한다. 철자 검사 및 교정 시스템에서는 일반적으로 잘못된 것을 바른 것으로 인식하는 문제와 맞는 것을 잘못된 것으로 인식하는 문제가 근간을 이룬다. 이 중에서 더 문제가 되는 것은 옳은 단어를 잘못 교정하는 문제점이다. 예를 들면, 옳은 복합명사인 '추가피해사실', '미라 주전투기'를 각각 '추가피해사실', '미라 주전투기

3) '꼬리말'이나 '접미명사', '접두명사'는 언어학에서 공식적으로 쓰는 용어가 아니라 본 철자/문법 검사기에서 어휘정보를 효율적으로 처리하기 위해 자체적으로 사용하는 명칭이다. 꼬리말 사전은 접미명사 범주에 속하는 단어들을 모아 놓은 것이다. 접미명사에는 '-사람', '-땅', '-령', '-어와 '-말' 등이 속하는데 여기서 '-사람', '-땅', '-말' 등은 접미사처럼 다른 명사 뒤에 붙여서 쓰지만 독립적인 단어로 홀로 쓰기도 하므로 엄격하게 접미사는 아니다. 따라서 접미명사라는 명칭을 쓴다.

접두명사도 같은 맥락으로 쓴다. 이와 같이 접미명사와 접두명사를 따로 처리하는 것은 파생어나 복합명사의 의미 결합 제약 규칙을 만들거나 띄어쓰기를 교정하는 데에 효과적이기 때문이다.

4) 하위범주화란 어떤 특정한 기준에 따라 분류한 하나의 범주에 속하는 요소들을 더 세밀한 분류기준에 따라 세분하는 것을 말한다.



〈그림 6〉 복합명사 결합제약 규칙베이스의 구성

가로 잘못 교정해 주는 예가 복합명사에서는 자주 나타난다. 왜냐하면 우리말 복합명사는 여러 단어가 병렬구조를 이루며 길게 이어져 나타나고 이 구성요소들 사이에는 하나의 문장처럼 어휘·의미적 관계가 존재하기 때문에 결합제약을 잘못 적용할 때 이와 같은 오류를 피할 수 없다. 한국어 복합명사에서 자주 발견되는 오류를 보면 대표적인 예를 세 가지로 들 수 있다[11].

발음이 유사하여 '공황극복'을 '공항극복'으로 쓰거나 '화재보험'을 '화세보험'으로 잘못 쓸 때가 종종 있다. 글자판에서 위치가 비슷하여 입력 오류를 범하는 예도 빈번하다. '나'와 'o'의 인접한 자판 위치로 인하여 '해상무역'을 '해산무역'으로 쓰는 예가 이에 해당한다. 사람들이 가장 자주 틀리는 띄어쓰기 오류도 빼놓을 수 없다. '공급자 수'를 '공급자수'로 쓰거나 '호주산 소고기'를 '호주산소고기'로 잘못 띄어쓰는 예를 들 수 있는데 이러한 예는 잘못 띄어 썼을 때 그 단어의 의미가 완전히 달라지기 때문에 매우 주의해야 한다.

3.2.2.1. 복합명사 결합제약 규칙베이스

복합 명사 결합 기능을 제약하는 방법으로는 형태적 제약방법과 의미적 제약방법을 이용하여 첫 번째 명사의 왼쪽과 오른쪽에 올 수 있거나 올 수 없는 명사들의 의미·형태적 특성과 명사가 나타나는 분포(distribution)에 따라 분류하여 규칙베이스화하였다.

가. 의미 결합정보의 표제어 선택

여기서 표제어는 주로 복합명사 결합오류를 일으킬

수 있는 단어이다. 표제어의 특성은 많이 사용되는 단어, 종성이 없는 음절로 구성된 명사나 잘못 쓸 가능성이 높은 단어, 위치적 제약을 가지는 단어들이다.

나. 결합제약 검증방향

의미 검사 후보 명사를 중심으로 왼쪽에 오는 명사에 제약을 적용하는지, 오른쪽에 오는 구성요소에 적용하는지를 표시한다. 왼쪽 우선 연관(Left Association) 비율이 70% 정도를 차지한다[13]. 또한 명사가 앞에 올 때와 뒤에 올 때 의미가 달라지는 예가 있으므로 순서(결합방향)를 고려하여야 한다.

다. 언어관계 (결합 관계)

표제어와 결합한 명사가 특정 단어와 결합할 수 있는 언어관계(collocation relation)에 있는지, 특정 단어와 결합할 수 없는 언어 오류 관계(anti-collocation relation)에 놓여 있는지 여부로 단어 간에 결합 가능 규칙인지 아니면 결합 불가능 규칙인지를 결정한다. 결합방향은 앞 뒤 즉, 좌 우 방향으로 결정되고, 결합이 가능한 지 불가능한 지에 대한 결정은 언어관계와 언어 오류 관계를 따르는가에 달려있다. 또한 '결합 가능'은 언어 관계를 나타내고 '결합 불가능'은 언어 오류 관계를 나타낸다.

라. 결합 제약 규칙

언어관계 정보는 어휘정보나 범주화한 의미정보(예: 사람, 동물, 식물, '하다'와 결합할 수 있는 명사 등), 개별단어를 이용하여 복합명사의 결합을 제약하는 규칙을 나타낸다. 규칙베이스에는 4,316개의 표제

〈표 4〉 규칙베이스의 분류정보 예

범주정보		설명	예
어휘· 의미적 분류 기준	사람명사	사람을 나타내는 한자어, 순수명사 등	어머니, 아버지, 경찰관
	동물명사	포유류, 조류, 파충류, 균류, 곤충	호랑이, 사슴, 돼지, 말
	식물명사	나무, 꽃, 채소 등	장미, 소나무, 배추
	지역명사	국가명, 도시명 등	한국, 일본, 서울, 부산
통사분류	'하다' 명사	'하다'와 결합할 수 있는 명사로 동작성 / 상태성을 표시	청구, 판매, 할인, 취소, 연기
어휘분류	외래어명사	외래어	프로, 아노미 등

어와 114개의 통사·의미·어휘적 하위범주정보를 가지고 명사 간의 의미적 결합을 제약한다. 범주정보는 명사의 형태적 특성과 의미에 따라 복합명사의 결합을 제약하려고 분류했다. 아래 표는 하위범주화한 정보의 일부를 소개한다.

마. 대치어 생성규칙

대치조건은 언어관계가 성립하지 않을 때 대치어 생성방법을 설명한다. 검사모듈이 의미적 언어관계를 검사한 후 언어관계가 성립하지 않으면 규칙에 정의된 대치 방법으로 대치어를 생성한다. 대치 방법으로는 단순 제약, 통사·의미 정보에 기반을 둔 단어 간 띄어쓰기, 의미정보에 기반을 둔 음절이나 단어 대치, 이 세 가지이다. 우선, 형태적 단순 제약 방법은 형태적으로 복합명사가 될 수 없는 단순한 명사 간의 결합에 기반하여 맞춤법 검사·교정기의 교정 모듈을 이용하여 교정한다. 예를 들면, '기계화영농'을 '기계화 영농'으로 교정한다. 다음은 통사·의미 정보에 기반을 둔 단어 간 띄어쓰기 방법으로서 오류일 가능성이 높은 명사가 포함된 복합 명사는 문체가 되는 이 명사를 중심으로 띄어쓰기를 하여 교정한다. 예를 들면, '대통령취임이전'에 대해 '대통령취임 이전'으로 띄어쓰기를 한 대치어를 제시한다. 세 번째는 의미정보에 기반을 둔 음절이나 단어를 대치하는 방법이다. 오류일 가능성이 높은 명사가 복합명사를 이룰 때 특정 음절이나 단어를 대치하여 의미를 올바르게 만든다. 예를 들면, '공부하녀'를 '공부하여'로 음절을 대치하여 의미적으로 결합할 수 있는 단어로 바꾸어준다.

3.2.2.2. 결합제약을 이용한 복합명사 교정

어떤 어절이 복합명사로 분석되거나 규칙화한 명사

가 복합명사로 생성되면 '복합명사 의미 분석 모듈'이 호출된다. 현재 어절의 구성요소에 특정단어의 의미상 언어관계를 검사할 필요가 있다고 판단되면 다음은 복합명사 결합제약 검사모듈을 호출한다. '복합명사 의미 분석 모듈'은 규칙베이스에서 의미 검사 후 보 명사의 결합 제약규칙을 찾아 좌·우 명사와 언어관계를 검사한다. 규칙베이스에는 각 명사에 대한 언어관계 정보, 검증방향, 결합 제약 규칙과 대치어 생성 규칙이 있었다. 이 때 언어 관계가 성립하면 복합명사일 가능성이 있음을 알리고 성립하지 않으면 규칙에 정의되어 있는 대치규칙을 적용하여 대치어를 생성한다. 주로 검사단어를 중심으로 언어 오류 관계에 있는 단어를 찾아서 언어 관계에 있는 단어로 대치어를 생성해 주는 방법을 취한다.

철자 검사기에서 의미적으로 처리하는 부분은 접미사, 접두사와 결합하여 단어를 이루는 파생명사와 복합명사였다. 구문이나 문장 단위로 검사하고 교정하는 문법 검사기에서 의미 처리는 어떻게 이루어지는지를 이제 알아보자.

가. 의미적 결합제약에 따른 띄어쓰기 대치

결합제약: '하다'명사 + '-자', '-인', '-가가 붙은 사람 명사 + '수'

-> '-가 수', '-인 수', '-자 수'

예) 사업가수 -> 사업가 수 (예외, 노동가수)

전문인수 -> 전문인 수

출국자수 -> 출국자 수

노동자수 -> 노동자 수 (예외, 동양자수)

결합제약: 식물명사+'의향' -> '~의 향'

예) 소나무의향 -> 소나무의 향

나. 의미적 결합제약에 따른 단어 음절 대체

결합제약: 하다명사+'하구' -> '하고' 예) 정리하구 -> 정리하고
결합제약: 하다명사+'회수' -> '횃수' 예) 납부회수 -> 납부횃수 (예외, 분할회수, 투자회수)
결합제약: 사람명사+'이식' -> '의식' 예) 학생이식 -> 학생의식

3.3. 문법 검사기

문법 검사기에서 처리하는 오류는 크게 문장구조 오류와 의미 오류 및 문체 오류로 분류할 수 있다. 우리말 문법은 매우 유연하여 필수적인 문장성분을 생략해도 문장이 이루어지기 때문에 문장의 뜻이 모호해질 때가 많으며 세부적으로 분석하지 않으면 오류를 찾아내기가 어려울 때가 흔히 있다. 의미 오류란 단어 간에 비슷한 발음이나 의미 때문에 혼동하여 잘못 쓰는 예이다. 문체 오류란 문장의 내용을 전달하는 표현 방법에서 어색하거나 불필요한 군더더기가 있는 겹말 오류, 상투적인 표현이나 우리말에 없는 영어, 일본어 번역체, 존칭 표현 오류 등을 이른다.

본 문법 검사기에서 이러한 다양한 문장 단위의 오류를 검사하고 교정하기 위해 의존문법에 바탕을 두

고 서술어와 논항 사이의 격정보를 분석하여 서술어를 하위범주화하고 서술어와 논항 사이의 연어정보와 체언의 연어정보(collocation)를 이용하여 문장을 부분적으로 분석하여 오류를 검사하고 교정하는 방법을 택한다.

3.3.1. 오류 유형

한국어 문법 검사기에서 처리하는 의미·문체 오류는 철자 오류와는 달리 어법상 완전히 틀린 말이 아니라 어색하거나 우리말의 고유성을 제대로 나타내지 못하는 표현을 처리하기 때문에 제시한 대체어의 사용을 권장하는 도움말을 제시한다.

<표 5>는 위에서 나열한 오류 유형 중에서 우리가 가장 자주 범하는 몇 가지 오류를 골라서 원인을 분석한 내용이다.

이 중에서 문체 오류는 주로 관용어구 표현처럼 하나의 패턴으로 묶어서 처리하는 예가 많다. 따라서 본 논문에서는 의미오류와 문장구조오류에 연구 초점을 두었다.

3.3.2. 문장구조 오류 교정

문장구조 오류는 필수적으로 들어가야 하는 문장 요소가 생략되거나 또는 자동사나 형용사가 목적어를 취할 때처럼 불필요한 문장 요소가 삽입되는 오류를 의미한다. 용언은 그 앞 성분 또는 뒤 성분과 결합관계

<표 5> 문법 검사기에서 처리하는 오류의 예

유형	오류 원인	오류 예	오류 교정
문장 구조 오류	여러 가지 문법적 의미로 해석할 수 있는 우리말 조사 '의'의 사용으로 인한 오류	나의 살던 고향은(x)	내가 살던 고향은(o)
	자동사/타동사 오류: '다치다'는 타동사로서 신체기관 명사를 목적으로 취하는 동사이다.	이마가 다치다(x)	이마를 다치다(o)
	자주 혼동하는 자동사/타동사: '가다'는 자동사로서 장소명사를 부사어로 취하는 동사이다.	철수는 학교를 간다(x)	철수는 학교에 간다(o)
의미 오류	유사한 의미나 발음 때문에 범하는 오류	통신사용료를 결제하다(x)	통신사용료를 결제하다(o)
	단어의 품사 혼동으로 범하는 띄어쓰기 오류	총알 수발을 쏘다(x)	총알 수발을 쏘다(o)
문체 오류	번역투 문장 오류: 일본어 번역투 영향	그 사실을 고려에 넣고(x)	그 사실을 고려하고(o)
	존칭 표현 오류: 전화를 건 '사람'이 아니라 사물인 '전화'를 잘못 높인 예	전화가 오셨습니다(x)	전화가 왔습니다(o)

를 통해 의미나 문법 관계가 제대로 드러난다. 자동사와 타동사의 가름은 구문론적 환경 제약에 따른다. 이

- 논파 100 3 0 2
- 논평 41 3 0 2
- 논간 100 3 0 3
-
- 논경 100 3 0 1
- 논담 41 3 0 1
- 논략 46 3 0 2
- 논사일 102 3 1
- 논사 41 3 0 4
- 논성 41 3 0 1
- 논작 100 3 0 1
- 논지거리 102 3 0 1
- 논축 46 3 0 2
- 논뉘름 100 3 0 1

(그림 7) 자·타동사 분류

연구에서는 동사를 자동사 타동사로 하위범주화하여 동사의 구문론적 환경에 오류가 있는지 검사한다.

문장구조에 기인한 오류를 교정하기 위해 자동사와 타동사, 자동사인 동시에 타동사인 동사를 나누었다. 현재 사전에서 동사로 분류되어 있는 총 16,912개 정보를 나눈 결과 11,562개 자동사, 3,128개 타

동사와 2,222개 자타동사를 추출하였다.

[그림 7]은 자·타동사 분류 자료의 일부를 발췌한 것이다. 정보가 붙은 순서는 표제어, 일반 사전 정보, 자·타동사정보(두 통사자질을 모두 갖춘 동사), 자동사(1)인지 타동사(2)인지를 구분하는 정보 순이다.

철수가 올해 시험을 합격하였다. (x)
 -> 철수가 올해 시험에 합격하였다. (o)
 자·타동사 규칙: 타동사는 목적격 조사 '-을/-를'을 반드시 동반한다. 그러나 '합격하다'는 자동사이므로 이 문장은 규칙에 위배되므로 오류로 판정된다.

자동사와 타동사를 잘못 쓴 다음과 같은 문장에서 자·타동사 규칙에 기반하여 오류를 발견하여 교정한다. 이러한 문장구조 오류의 원인 중에 하나가 영어와 같은 외국어를 우리말로 번역할 때 각 나라 언어마다 서로 다른 동사의 통사자질을 고려하지 않고 직역을

<표 6> 자주 혼동하는 자동사와 타동사 구문 오류의 예

자동사	자동사를 타동사로 잘못 쓴 예	
	오류 구문	교정한 예
가다	학교를 가다	학교에 가다
가입하다	동호회를 가입하다	동호회에 가입하다
복학하다	대학을 복학하다	대학에 복학하다
성공하다	사업을 성공하다	사업에 성공하다
숙달하다	방법을 숙달하다	방법에 숙달하다
실패하다	결혼을 실패하다	결혼에 실패하다
임하다	전쟁을 임하다	전쟁에 임하다
적응하다	새 생활을 적응하다	새 생활에 적응하다
접근하다	서양 문화를 접근하다	서양 문화에 접근하다
접어들다	가을을 접어들다	가을에 접어들다
진출하다	국외 시장을 진출하다	국외 시장에 진출하다
착륙하다	달을 착륙하다	달에 착륙하다
출강하다	대학을 출강하다	대학에 출강하다
합격하다	사법고시를 합격하다	사법고시에 합격하다
타동사	타동사를 자동사로 잘못 쓴 예	
	오류 구문	교정한 예
다치다	무릎이 다치다	무릎을 다치다
멤돌다	주위로멤돌다	주위를멤돌다

〈표 7〉 문법 검사기의 의미처리 규칙베이스

검사단어	의미 규칙	실제 교정 예
영키다	-검색할 대상 단어 -대치 단어	{주어: 액체(피, 기름, 물, 우유 따위)} -> 영키다
엷히다	-대치어 수 -교정 유형: 단순대치, 대치단어 앞 뒤 분리 -현재 단어 제약 조건: 명사+조사, 용언+어미	{주어: 실, 줄 종류} -> 엷히다
영기다	-검색 방향: 좌, 우, 양측 -체언의 의미 분류 정보: 격 정보와 의미 정보	{주어: 상황, 일, 사건, 관계 따위} -> 엷히다

하는 바람에 우리말에서까지 잘못 쓰고 있는 점이다. 따라서 자동사와 타동사의 분류작업은 번역투 문체 오류를 교정하는 데에도 큰 역할을 하였다. 특이할 점은 우리말 자동사를 타동사로 잘못 쓰고 있는 구문 패턴은 62개를 발견했지만 타동사를 자동사로 잘못 쓰는 예는 그 수가 매우 미미하다는 것이다.

3.3.3. 서술어/논항 관계와 의미 오류 교정

문장의 의미는 문장에 있는 단어의 의미와 의존관계에 있는 단어 간의 의미관계의 총합이다. 의미해석의 목적은 이와 같은 의미표현을 얻기 위하여 두 가지 타입의 중의성을 해소하는 것이다. 하나는 문장 내에 있는 다의어에 관한 의미적 중의성 해소, 다른 하나는 의존관계에 있는 두 단어 사이의 의미관계에 관한 의미적 해소이다.

본 논문에서는 특히 두 단어 사이의 의존관계를 통해 의미적으로 처리하는 방법에 대해 알아본다. 한국어 문법 검사기에서 사용하는 의미 처리 방법은 크게 언어관계를 이용한 부분 문장 분석과 의미 분석 방법과 어휘와 구조의 중의성 해소 규칙이다. 이러한 의미 처리 방법을 고려하여 본 검사기에서 구축한 의미 처리 규칙베이스는 다음과 같다.

문장에서 가장 중심요소인 용언의 의미정보 처리를 하려면 용언의 통사적 자질을 기반으로 한 자동사, 타동사 분류와 논항의 의미정보를 고려한 용언의 통사·의미적 자질을 근거로 하위범주화한다. 서술어와 논항 형식으로 표현하는 것은 각 문장을 하나의 서술어와 그에 따른 논항이 구성하는 서술적인 관계로 표현하는 것이다(14).

먼저 자동사·타동사 분류 작업을 통하여 통사적 자질이 변별 기능을 충분히 할 때에는 논항의 의미정보는 필요하지 않지만, 문맥 속에서 서술어의 앞과 뒤에 있는 논항의 정보를 보아야만 변별할 수 있는

용언이 있다.

- ① 통사자질로 구별할 수 있는 동사
 - { 안다: 타동사 + 목적어(대격: -을/-를)
 않다: 자동사 + 부사어(장소격: -에)
 그외 '낫다/날다', '붓다/뿜다' 등 }
- ② 논항의 의미자질로 구별할 수 있는 동사
 - { '다리다/달이다', '붙이다/부치다',
 '영기다/영키다', '저리다/절이다' 등 }

용언의 논항이 되는 명사를 의미적으로 분류하는 작업은 곧 국어사전, 단어의 용례에 관한 인간의 상식, 명사의 isa 계층구조를 포함하는 지식베이스 구축을 필요로 하기 때문에 간단한 과제가 아니다.

특히, isa 계층구조는 각 명사의 상위 개념을 의미망(semantic network) 형식으로 만들어 놓은 것이다. 이러한 계층구조는 인간의 개념 인식 과정을 컴퓨터에게 부여하는 하나의 방법이다. 한국어 단어의 의미망(Wordnet) 조직에 대한 연구는 현재 진행되지 않고 있는 실정이다. 단어의 의미망 조직(Network)이라는 개념이 광범위하게 확산되었지만, 이를 실제 연구 사업으로 추진하는 수준에는 미치지 못한 것이다. 다만 새롭게 추진되고 있는 세종 계획의 전자 사전 집필 형식을 살펴보면, 단어의 의미망(Wordnet)에서 기본적으로 사용되고 있는 의미론적 관계를 포괄하고 있는 것을 확인할 수 있다[고려대학교 한국학 교육연구단 자료(<http://ikc.korea.ac.kr/~eriks2k>)].

본 문법 검사기에서 명사의 의미적 분류작업은 체계적으로 광범위하게 이루어지지는 않았다. 우리가 자주 틀리는 오류의 쌍을 모아서 일반적인 규칙을 만들어내고 추가하는 단계에서 필수적인 구성요소들을

의미정보 분류 작업을 한다. 그 중에서 가장 중요한 동사의 하위범주화작업에서 동사의 의미적 쓰임을 결정짓는 논항의 의미정보 분류하고 오류쌍에 따라 변별 단계를 정할 수 있다. 예를 들면, 동사 '삭히다'는 주류(술) 정보를 가진 명사를 목적어로 취하며, 대치 단어인 '삭이다'는 일반적인 음식물 정보나 감정 명사인 '분'을 목적어로 가진다. 이 경우 음식물 정보의 하위 범주는 세분되어야 한다. 반면 '영키다'와 '영기다'를 검색하는 데에는 상위 범주인 '액체 정보'가 변별 자질로서 충분하다.

〈표 8〉 서술어와 논항관계에 따른 의미정보 변별 단계

서술어/논항 관계	논항의 의미적 변별자질
(술, 감주, 막걸리, 등)을 삭히다 (음식물, 분)을 삭이다	{±주류} / {±음식물} {±감정}
(실, 줄, 머리카락, 등)이 영키다 (피, 우유, 기름, 등)이 영기다	{±액체}

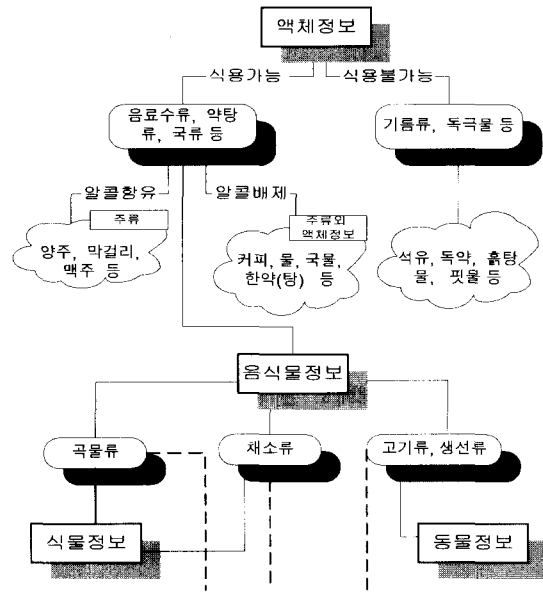
의미 정보 분류의 기준은 서술어의 통사·의미 자질에 의거한다. '삭히다'와 '삭이다'의 예를 보자. 먼저 상위 범주인 '음식물'에서 하위 범주인 '주류'를 세분한다. 반면, 현재로서는 추상 명사 '분'처럼 분류하기 까다로운 정보를 하나의 상위 범주 '감정' 속에 포함시켜도 변별자질로서는 충분하다. '영키다'와 '영기다'의 경우를 보자. '영기다'의 논항의 의미 정보가 일반적인 상위 범주 '액체 정보'에 등록되어 있으면, '영키다'의 논항이 어떤 범주에 속하는가와는 무관하게 두 동사는 구별된다. 서술어의 의미에 따른 분류의 장점은 이러한 작업의 경제성에 있다.

하나의 단어는 여러 가지 상위 범주를 가질 수 있다. 예를 들면, 음료수류, 국물류 등은 액체뿐만 아니라 음식물에도 속한다. 또한 채소류나 곡물류는 음식물과 식물을 상위개념으로 취할 수 있다. 즉, 기본개념은 여러 최상위 개념에 의해서 교차분류(cross-classification)가 허용된다[15, 21].

3.3.4. 부분 문장 분석과 언어관계

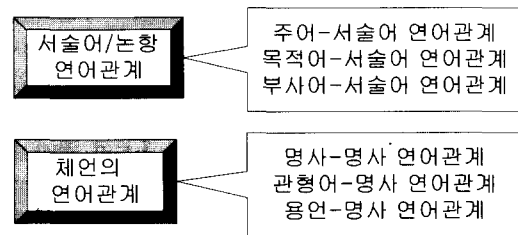
언어관계⁵⁾도 지금까지 서술한 동사의 격틀정보, 서

5) 언어관계(collocation)란 특정한 주어와 동사 또는 목적어와 동사처럼 두 개의 어휘성분이 자주 어울려 나타나면서 이들의 어울림을 근거로 그 정확한 의미를 쉽게 파악할 수 있는 어휘쌍이다.



(그림 8) 명사의 의미적 분류의 예

술어-논항의 관계를 표현하는 이론적 근거를 이루는 의존문법과 무관하지 않다. 비교적 좁은 영역의 주변 단어들이 주어질 때에도 단어의 중의성을 해소할 수 있다는 사실이 언어심리학적 실험을 통해 밝혀졌다. 이 사실은 단어의 의미 중의성을 해소하는데 있어서, 그 단어를 둘러싼 좁은 범위의 주변 단어들만 잘 활용하여 좋은 성능을 얻을 수 있다는 것을 간접적으로 보여준다. 단어 의미 중의성에 관한 연구들에서는 이러한 사실을 국소문맥(local context) 또는 언어정보(collocation)라는 용어를 사용하여 다양한 방법으로 이용하였다[16]. 국소문맥은 지엽적인 의존관계(local dependency)[17]와 통하는 말로 서술어-논항 관계에서 성립하는 동사-주어, 동사-목적어 사이의 관계, 수식어-피수식어 관계처럼 바로 인접한 문맥에서 상호



(그림 9) 문법 검사기에서 쓰는 언어정보

〈표 9〉 격들정보로 서술어의 의미규칙을 표현한 예

주어	목적어	부사어	동사 교정
[+액체]이/가	없음	임의 요소	엇기다(X) → 엇기다(O)
[+신체]이/가	없음	임의 요소	질이다(X) → 저리다(O)
[+사람] [+동물]이/가	[+음료수]을/를	임의 요소	들이키다(X) → 들이켜다(O)
[+사람]이/가	[+신체 기관]을/를	임의 요소	의식하다(X) → 이식하다(O)
[+사람]이/가	[+공간], [+방향]을/를	임의 요소	가르치다(X) → 가리키다(O)
[+사람]이/가	[+학문], [+기술], [+스포츠]을/를	임의 요소	가리키다(X) → 가르치다(O)
[+사람]이/가	없음	[+공간]에	들리다(X) → 들르다(O)
[+사람]이/가	[+고기류, 생선류]를	[+연료류: 불]에	그을다(X) → 그슬다(O)

한정할 수 있는 관계를 이르는 말이다.

우리말에서 언어관계가 성립할 수 있는 어휘정보를 살펴보면 〈그림 9〉에서 보여주는 것처럼 서술어-논항 관계와 체언의 언어관계로 크게 분류가 된다.

이 중 문법 검사기에서 주로 처리하는 대상은 서술어-논항의 언어관계이다. 서술어와 논항의 역할을 하는 명사 사이의 의존관계를 총괄적으로 서술하기 위해 격들정보를 이용하면 여기에서 나타나는 의미표현이 바로 언어 정보임을 알 수 있다.

격들정보는 통사구조와 의미구조를 연결해주는 역할을 담당한다. 이 정보는 두 가지 목적을 보인다. 첫째, 주어진 한 구문에서 어떤 의미관계를 가질 것인지를 결정함에 있어서 통사구조 정보를 사용하는 근거를 제공한다. 둘째, 현재 의미문맥에 적합하지 않은 단어에 통사적인 서술성을 제한하고 적절하지 않은 의미관계를 버림으로써 정확하지 않은 해석을 제거하려는 비교적 단순한 원리(메카니즘)이다(18).

〈표 9〉는 자주 틀리는 동사의 오류쌍에 대하여 격들정보를 이용하여 의미규칙으로 표현한 예이다.

〈표 9〉를 살펴보면 언어 정보를 역으로 이용하여 오류를 검사해 낼 수 있다. 다시 말해, 검사단어와 언어 오류 관계(anti-collocation)에 있는 단어를 찾아내서 교정해 준다. 이 방법은 의존 관계 허용 범위를 넓혀서 언어 오류 단어를 찾을 수 있는 범위를 확대하고, 의존 관계를 설정할 때 소요되는 시간을 줄여준다. 그러나 분석 범위가 너무 확대될 때 문법 검사기의 속도 저하와 검사 오류가 생길 수 있다. 이 문제를 해결하려고 문법 검사기가 검증할 수 있는 어절 수를 제한하거나 문장 분석을 종료하는 조건(Stop-Condition)을 둔다(19, 20).

문장 분석 종료 조건(Stop-Condition)은 용언의 종결형어미나 문장 종결 부호(!, ?, . 등)가 올 때 해당한다. 그러나 검사 단어와 언어 오류 단어는 대체로 동일한 주어나 목적어나 서술어를 가지는 문장 범위 안에 있다. 이러한 특징을 고려하여 문법 검사기에서는 검사 단어를 기준으로 문장 분석을 하다가 문장의 서술어가 동일한 주어나 목적어를 가질 수 있는 연결형어미 '-며'나 '-고' 등이 오면 문장 분석을 끝낼 수 있도록 조건을 만들었다.

철수는 서산 쪽을 보고 손으로 가르쳤다 (X)
 (언어오류단어) (검사단어)
 => 철수는 서산 쪽을 보고 손으로 가리켰다 (O)
 검사단어 '가리키다'
 언어관계 단어: [+학문], [+기술], [+스포츠]
 언어 오류 관계 단어: [+공간], [+방향]

위에서 서술어 '가리키다'와 목적어 '쪽'은 서로 언어 오류 관계에 있는 단어들이다.

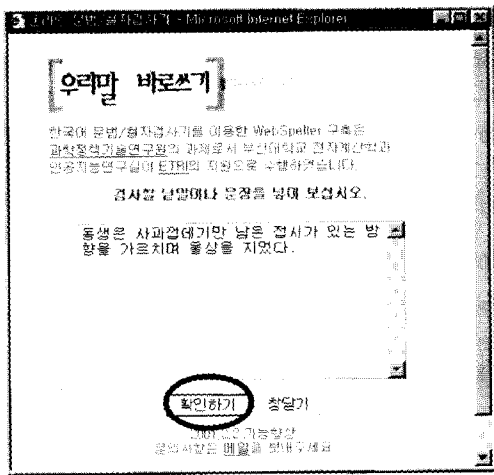
먼저, 검사단어 '가리키다'를 중심으로 앞 방향 분석을 하다가 '보고'가 나오면 한 문장 내에서 같은 목적어를 가질 수 있는 연결형 어미 형태인지를 확인한다. 필수 문장 요소인 목적어나 서술어가 아직 발견되지 않았고 동일 목적어를 가질 수 있는 서술어이므로 계속 문장 분석을 한다. 그러나 그 다음 단어 '쪽'이 '가리키다'의 언어 오류 단어이므로 '가리키다'로 교정한다.

담론이나 주어와 같은 병렬문 구조에서는 부분문장 분석과 언어정보를 이용하여 의미 오류나 문체 오류를 큰 무리 없이 처리할 수 있지만 더 복잡한 문장을

처리하는데에는 여러 가지 한계가 있다. 우리는 이에 대해 결론에서 구체적으로 언급하겠다. 이 외에도 문장에서 단어의 중의성을 해소하는 처리방법이 있으나 본 논문에서는 특히 단어와 단어 사이의 의존관계의 애매성 해소에 초점을 두었다.

4. 학습시스템에서 한국어 철자/문법 검사기 활용

이 논문에서 지금까지 소개한 우리말의 다양한 오류에 대한 의미처리는 한국어 철자/문법 검사기의 성능을 더욱 향상시켜 한국어 학습시스템 '우리말 배움터'(http://urimal.cs.pusan.ac.kr)와 연동을 통해 웹 환경에서 더 효율적으로 서비스하기 위한 것이다.



(그림 11) 한국어 철자/문법 검사기 실행

그리하여 한국어 철자/문법 검사기가 단순한 철자 오류의 범위를 뛰어넘어 문장 단위에서 문맥상 틀린 단어를 찾아서 고쳐주는 기능을 제대로 수행하여 교육적인 효과를 거둘 수 있도록 한다.

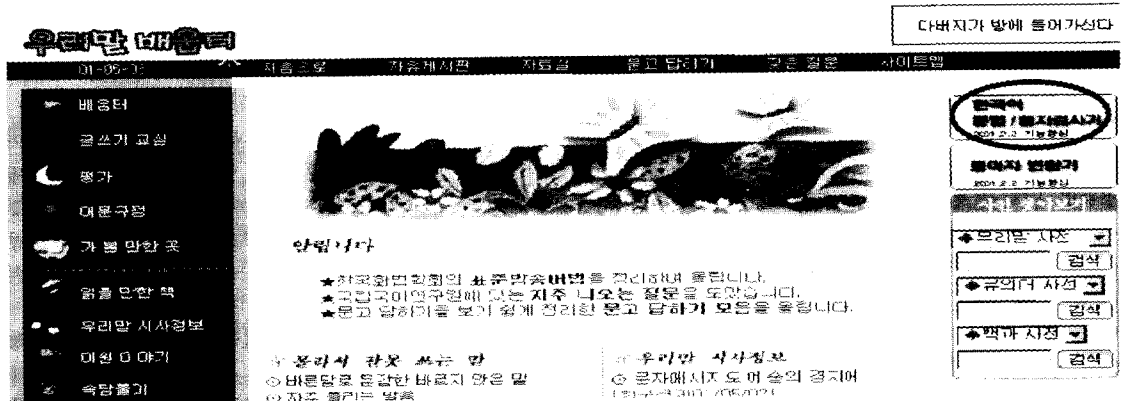
4.1. 한국어 철자/문법 검사기 실행

한국어 철자/문법 검사기는 오류를 검사하여 대처어를 생성해 주고 도움말을 제시하는 기본 구조를 가지고 있기 때문에 각 오류에 해당하는 도움말을 다양하게 갖추고 있다. 오류의 수만큼이나 대량으로 축적된 도움말 정보는 한국어 학습시스템에 백과사전식 지식을 제공해 줌으로써 한국어 학습내용으로 활용하기에 손색이 없다. 이제 한국어 철자/문법 검사기는 인터넷에 접속만 하면 누구든지 자신의 어휘력이나 문장력을 점검해 보고 관련 학습을 할 수 있게 되었다.

온라인 한국어 철자/문법 검사기 입력창에 문장을 새로 입력하거나 사용자가 이미 작성한 문서를 넣는다. 입력 창에서 받아들인 내용은 webspeller.dll로 넘어가서 다음과 같이 처리된다.

```
<form name=form action=webspeller/webspeller.dll?check*
<textarea name=text>/textarea>
<input type=submit name=철자 검사>
</form>
```

철자검사는 사용자가 입력한 문장을 분석하여 오류를 검사하고 대처어와 도움말을 생성하고 이 오류와 관련이 있는 학습페이지를 연결해 준다.



(그림 10) 우리말 배움터 학습시스템과 한국어 철자/문법 검사기

학습자가 똑같은 오류를 반복하지 않도록 하기 위해 오류에 해당하는 학습내용을 연결해주는 매개체는 한국어 철자/문법 검사기의 도움말 사진이다. 아래에서 보는 바와 같이 오류 유형에 따라 도움말의 번호 구간이 정해져 있고 각 도움말 번호 구간에 관련 학습페이지의 URL이 연결되어 있다.

4.2 오류 유형에 따른 자율 학습

학습자가 '우리말 배움터' 학습시스템에 연결되어 있는 한국어 철자검사기를 클릭하여 문장을 입력하거나 자신이 쓴 문장을 붙여 넣은 후의 학습 줄거리를 따라가 보면 다음과 같다.

① 학습자가 쓴 문장에서 문장성분(주어, 목적어, 용언)이 서로 일치하지 않는다.

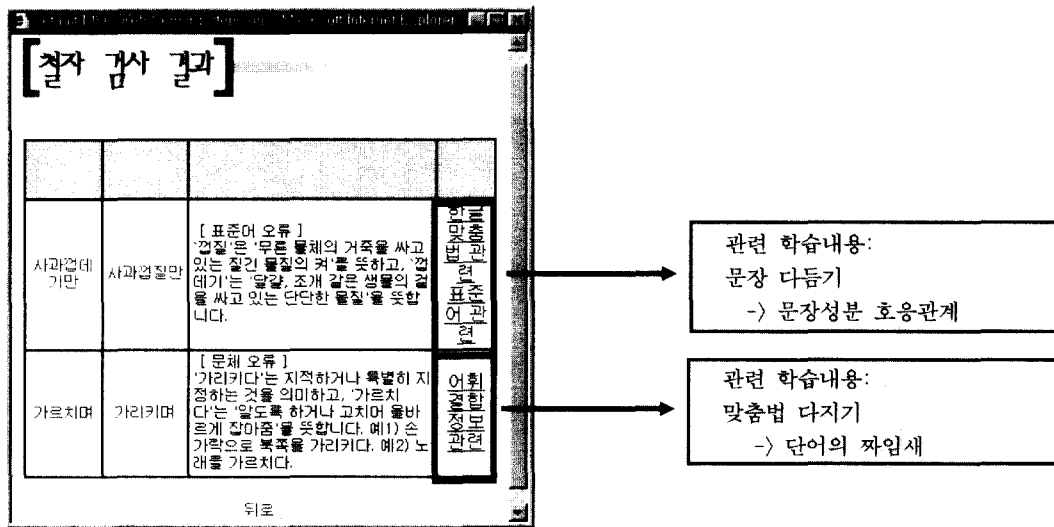
② 한국어 문법 검사기는 구문분석을 통해 문장의 통사적 오류를 발견한다.

③ 오류에 대한 대처어를 생성하고 해당 도움말을 제시한다.

④ 도움말과 관련이 있는 학습 문서를 도움말 번호 구간과 연결되어 있는 URL을 가지고 연결한다.

⑤ 학습자는 학습시스템에서 있는 학습 과정을 필요한 내용을 익힌다.

앞서 입력한 문장을 철자 검사기로 실행한 결과, '사과곶대기'와 '가르쳤다'가 오류로 판정되었다.



(그림 12) 철자 검사 결과와 관련 학습

```

CString CWebSpellerExtension::GetSite(SPCHKRES_FORMAT lpSpRes)
{
    int E_code = lpSpRes.nErrCode;
    if (E_code > 0 && E_code <= 999) // USER define
        return _T("");
    else if (E_code >= 1000 && E_code <= 1999) // 맞춤법
        return
        _T("<a href='\"../../sitemanager/anchor_manage/url_search.asp?key=1010100'>한글 맞춤법 관련");
    else if (E_code >= 2000 && E_code <= 3999) // 표준어
        return
        _T("<a href='\"../../sitemanager/anchor_manage/url_search.asp?key=1020100'>표준어 관련");
    else if (E_code >= 4000 && E_code <= 4999) // 오용어
        return
        _T("<a href='\"../../sitemanager/anchor_manage/url_search.asp?key=1010100'>한글 맞춤법 관련<br>");
    _T("<a href='\"../../sitemanager/anchor_manage/url_search.asp?key=1020100'>표준어 관련");
    //(...)
}
    
```

우리말 배움터
01-05-16
처음으로
자유게시판
자료실
묻고 답하기
찾은 질문
사이트맵

배움터

- ★ 맞춤법 다지기
- ★ 어휘 다듬기
- ★ 문장 다듬기
- ★ 문단 다듬기

전체 검색

방문자 수 109637

단어의 짜임새

단어는 하나의 자립된 말로 된 단말어. 실질적 의미를 나타내는 중심 부분인 어근과 어근에 붙어 그 뜻을 제한하는 접사로 이루어진 파생어. 두 개 이상의 자립어가 결합한 복합어 등이 있습니다. 이와 같이 다양한 단어가 어떻게 형성되는지를 알아봅시다.

여러말 어어의 결합관계

체언과 조사 > 꽃에 물을 준다? 꽃에게 물을 준다?	어간과 어미 > 이 편지는 급히 부쳐야 돼. 우표도 반드시 붙이고
접미사 > 발을 깨끗이 치우고, 뽀건은 제자리에 반듯이 놓아라.	합성어와 접두사 > 왜 '술나무'를 '소나무'로 '바늘질'을 '바느질'로 쓰는가?
사이시옷 > 마룻마를 사람들은 뒹새가 너무 심해	준말의 형태 > 뉘러지 않은 생활에 드는 비용이 만만치 않다

(그림 13) 단어 짜임새에 대한 학습

우리말 배움터
01-05-16
처음으로
자유게시판
자료실
묻고 답하기
찾은 질문
사이트맵

배움터

- ★ 맞춤법 다지기
- ★ 어휘 다듬기
- ★ 문장 다듬기
- ★ 문장성분 호응관계
- ★ 수식어와 피수식어
- ★ 사동피동형
- ★ 번역투 표현
- ★ 문단 다듬기

문장성분 호응관계

문장 성분이란 한 문장을 구성하는 요소를 말합니다. 문장에서 가장 핵심이 되는 요소는 서술어이며 이 서술어의 어휘적 특징과 문법적 기능에 따라서 서로 다른 문장 성분을 선택합니다. 이 때 문장의 필수적인 주성분이 일치하지 않으면 완전한 문장이 될 수 없습니다

서술어와 주요 문장 성분 사이의 호응관계

서술어와 주어 > 그는 아이들과 더불어 운동장에서 공을 차고 있는 친구를 본다	서술어와 목적어 > 사람들은 즐거운 마음으로(일률?, 공 부를?)할때 능률어 오른다.
---	---

(그림 14) 문장성분 호응관계 학습 내용

'껌테기'는 '조개'와 같이 딱딱한 물질을 깔을 싸고 있는 것을 뜻하므로 과일류 명사와는 결합할 수 없다.

이와 같은 오류를 범한 사용자를 위하여 본 학습시스템에 '맞춤법 다지기'의 '단어의 짜임새' 부분을 공부할 수 있도록 연결하였다.

앞서 나온 예문에서 목적어 '방향'에 동사 '가르쳤다'를 쓴 사용자라면 동사 '가르치다'와 '가리키다'를 제대로 구별하지 못하기 때문에 이러한 오류를 범하는 것으로 유추할 수 있다. 이에 문장에서 필수적인 문장성분인 주어나 목적어와 서술어 사이의 호응관계에 대한 내용을 학습할 수 있도록 연결한다.

그러나 사용자가 원하는 학습내용을 더욱 상세하고 구체적으로 제공하려면 도움말 번호와 학습 사이트의 URL을 구분자로 하여 찾아 주는 방법만으로는 사용

자가 완전하게 만족할 수 있는 내용을 전달할 수 없다. 학습자가 범한 글쓰기 오류를 더 세밀하게 분석하여 도움말 번호가 아닌 핵심단어를 구분자로 하여 더 구체적인 학습 사이트를 지정해 줄 수 있는 작업을 향후 수행해야 한다.

5. 결론 및 향후 과제

한국어 철자/문법 검사기를 학습시스템에 연결한 목적은 학습자의 직접적인 학습참여를 유도하여 능동적인 학습을 할 수 있는 지능적인 시스템을 만들고자 하는 것이었다. 따라서 한국어 철자/문법 검사기가 형태소 분석 단계에서 가능한 단순한 철자 오류만을 해주는 정도의 성능만을 보인다면 국어 작문에 도움을 받으려고 이 시스템을 쓰는 사람들에게 쉽게 그

한계를 드러내는 결과를 초래할 것이다. 이 논문의 본론에서는 한국어 철자/문법 검사기의 의미처리를 주요 내용으로 다루었다. 그러나 한국어 철자/문법 검사기에서 아직도 의미적으로 해결하기 어려운 문장이 몇 가지 있다. 그 중 대표적인 예가 영어의 관계 대명사절과 같은 유형의 내포문이 삽입된 문장이다.

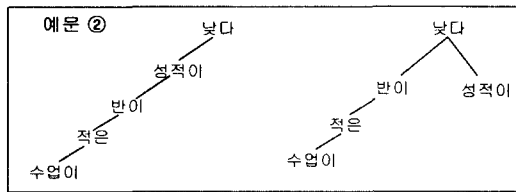
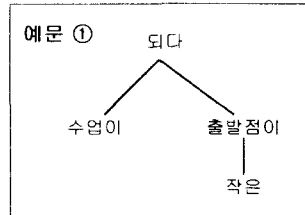
①수업만 바뀐다고 해서 학생들에게 창의성이 손쉽게 생겨나지는 않지만, 수업이 작은 출발점이 될 수는 있다. [19]

-> 수업이 적은 출발점이 될 수는 있다.(x)

②수업이 작은 반이 성적이 낮다

-> 수업이 적은 반이 성적이 낮다 (o)

(의미해석1: '수업이 적은 반의 성적이 낮다'
의미해석2: '수업이 적은 반에 성적이 낮다')



예문 ①에서 '작은'은 앞에 있는 주어 '수업'의 서술어가 아니라 바로 뒤에 오는 명사 '출발점'을 수식하는 말이므로 교정해서는 안된다. 그런가하면 예문 ②에서 '작은'은 '반'을 수식하는 말이 아니라 '수업'을 주어로 취하는 서술어이므로 '적은'으로 교정해야 한다. 이 두 문장에서 '작은'은 인접한 문맥 속에 '수업'과 '반'을 취하고 있기 때문에 현재 문법 검사기의 부분 문장 분석 방법으로는 해결하기가 어려운 문제이다.

이러한 중문이나 복문과 같은 문장 처리문제는 현재 학습시스템에서 '맞춤법 다지기'와 '어휘 다듬기' 학습보다는 수준이 더 높은 '문장과 문단 다듬기' 단계의 학습에 앞으로 연구할 과제를 남긴다.

한편, 웹기반 학습 관점에서 가상학습은 공간적·시간적 제한이 없다는 강점이 있지만 교실학습에서처럼 교사가 모든 학습과정을 통제할 수 없으므로 학습자가 주도적으로 학습자료를 적절하게 이용하여 학습을 진행해 나가겠다는 책임감을 가져야 한다. 그러다

보니 학습자가 스스로 학습하려는 굳은 의지가 없이는 꾸준하게 학습을 하지 못하고 쉽게 싫증을 느낄 수 있다. 이에 대한 해결책으로 다음과 같은 제안을 한다.

첫째, 잠재적인 학습자의 수준과 취향, 지식 정도, 연령, 관심 정도와 그에 맞는 학습내용을 구성하여 제공한다. 둘째, 다양한 학습 즐거리를 소개하여 사용자가 선택할 수 있게 한다. 셋째, 도움말 번호구간이 아닌 구체적인 핵심어를 매개체로 오류 유형에 일치하는 세부 학습내용을 연결해 준다. 끝으로, 학습자의 참여도나 학습진행 상황을 점검할 수 있는 지능적인 관리시스템을 개발한다. 이 네 가지 해결책은 앞으로 계속 연구해야 할 향후 과제이기도 하다.

참고 문헌

- [1] 유인환, 신수범, 이철현, 이태욱(1999). 자기주도적 학습을 위한 웹 기반 교육환경의 평가 기준 방안. 한국정보과학회 봄 학술발표논문집 vol.26 n.1, 697-699
- [2] 손병길(1999). 제7차 교육과정 실현을 위한 교육용 소프트웨어 개발·보급. 정보과학회지 특집 교육정보화 제17권 제9호 통권 제124호, 9월, 19-27
- [3] Hartley Stephen & Medley David(1996). Enhancing Teaching using the Internet: Report of the Working Group on The World Wide Web as an Interactive Teaching Resource. ACM, 218-228
- [4] 남현숙, 김상훈, 김지원, 권현주, 정유진, 권혁철 (2000). 한국어 철자/문법 검사기와 웹 기반 언어 학습시스템의 통합환경 구축. 한국인지과학회 2000 춘계학술대회 논문집, 연세대학교, 37-40
- [5] Rosenfeld Louis & Morville Peter(1998). Information Architecture for the World Wide Web, OREILLY, USA.
- [6] Reeves Thomas & Reeves Patricia. Effective Dimensions of Interactive Learning on the World Wide Web [번역본 <http://www.cnuc-e.ac.kr/~ydp>]
- [7] Manning Christopher & Schüze Hinrich(1999). Foundations of Statistical Natural Language Processing. MIT Press, London
- [8] 이지선, 전현경, 김남수, 이홍석(2000). WordNet

- 을 이용한 한국어 명사 의미지표 자동 구축, 정보과학회 학술대회 논문집, 대구 효성 카톨릭대, 333-335
- [9] 권혁철(1997), 한국어 철자 검사기와 문법 검사기, 정보과학회지 특집 국어공학 제15권 제10호, 10월, 24-34
- [10] 권혁철, 윤애선, 최준영(1991), 단일화 기반 의존 문법에 의한 자연언어 분석 기법, 한국정보과학회 봄 학술발표논문집
- [11] 원상연, 김수남, 김광영, 남현숙, 권혁철(1999), 한국어 문법 검사기에서 의미정보를 이용한 복합명사의 분석 제약, 제 11회 한글 및 한국어 정보처리 학술대회 논문집, 전북대학교, 288-293
- [12] 신수범, 유인환, 이철현(1999), 컴퓨터 교육 내용에 따른 원격 컴퓨터 교육 모형 탐색, 한국정보과학회, 가을 학술발표 논문집, 685-687
- [13] 강승식(1996), 한국어 형태소 분석을 위한 복합명사의 인식방법, 제 8회 한글 및 한국어 정보처리 학술대회 논문집, 241-245
- [14] Fuchs Catherine(1993), Linguistique et Traitements Automatiques des Langues, Hachette
- [15] 김현권(2000), EuroWordNet의 구성원리와 설계, 언어학 제27호, 한국언어학회, 12월, 145-177
- [16] 이승우, 이근배(2000), 국소문맥과 공기정보를 이용한 비교사 학습방식의 명사의미 중의성 해소, 정보과학회 논문지: 소프트웨어 및 응용, 제27권 제 7호, 6월, 769-783
- [17] Rastier François Cavazza Marc, Abeille Anne(1994), Pour l'Analyse Sémantique de la Linguistique à l'Informatique, Masson
- [18] Hendrix Gary(1986), Semantic Aspects of Translation, in Readings in Natural Language Processing, Morgan Kaufmann Publishers, Inc., 267-283
- [19] 소길자, 권혁철(2001), 어휘적 중의성 제거 규칙과 부분문장분석을 이용한 한국어 문법 검사기, 정보과학회논문지: 소프트웨어 및 응용, 제28권 제 3호, 3월, 305-315
- [20] Nam HyeonSook, So KilJa, Kim SuNam, Kwon Hyuk-Chul(1998), Resolving Lexical and Syntactic Ambiguities Using Heuristic Rules for a Korean Grammar Checker, The 4th ICCSI, 10월 26-31일, North Carolina USA, 454-457
- [21] Aitichison Jean & Gilchrist Alan, 전학병 옮김 (1991), 시소러스의 작성법, 산업기술정보원
- [22] 김광영, 남현숙, 박수호, 권혁철(2000), 한국어 철자 검사기의 교정기법 개선, 한글 및 한국어 정보 처리 학술대회 논문집, 성공회대, 89-94