

어휘적 중의성 제거 규칙과 부분 문장 분석을 이용한 한국어 문법 검사기

(A Korean Grammar Checker using Lexical Disambiguation
Rule and Partial Parsing)

소길자[†] 권혁철^{**}

(Kil-ja So) (Hyuck-chul Kwon)

요약 본 논문에서는 우리말 문서에 있는 오류를 어절 단위로 검증하는 철자 오류와 여러 어절을 분석해야 처리할 수 있는 문법 오류로 분류하였다. 문법 오류를 처리할 때 전체 문장 분석은 시간이 많이 소요되고 구현하기 어려우므로 대부분 부분 문장 분석 방법을 이용한다.

기존 연구에서 사용한 부분 문장 분석은 분석 어절에 어휘 중의성이 있을 때 문장 분석 종결 또는 과 분석 등의 오류가 발생한다. 본 논문에서는 문법 검사기에서 어휘 중의성 때문에 발생하는 문제점을 해결하는 방법으로 어휘 중의성 제거 규칙을 사용한다. 본 논문에서 구현한 어휘 중의성 제거 모듈은 코퍼스 데이터에서 얻은 경험적 규칙을 기반으로 한다. 이 경험적 규칙은 언어적 지식을 기반으로 한다.

Abstract In this paper, we classify the errors found in Korean documents into two types: One is spelling errors that can be detected by analyzing a word (or "eu-jul" in Korean term) and The other is grammatical errors that can be detected by analyzing multiple words. Most previous approaches use partial parsing method because full-sentence analysis is time consuming and difficult to implement. However, partial parsing method depend on a sub-system to predict accurately the correct part of speech for each word because of the limitations of coverage. In this paper, we propose a different approach that of previous words. The propose system applies disambiguation rules while detecting grammatical errors by partial parsing.

We implemented the system using heuristic rules based on the results obtained from the analysis of corpus data. The heuristic rules include the disambiguation and syntactico-semantic rules.

1. 서론

1.1 연구 배경

기존 철자 검사·교정기는 일반적으로 한 어절을 대상으로 검사를 하므로 문맥상 어울리지 않는 단어 때문에 생기는 오류는 찾지 못한다. 다수 어절에 걸쳐 검증해야 하는 오류를 정확히 교정하려면 구문분석을 행해야 한다[9]. 그러나 구문 분석을 하면 구문 오류를 정확

히 찾을 수는 있으나, 속도가 느리고 많은 기억 공간이 필요하므로 효율적이지 못하다. 이를 해결하려고 부산대학교에서 개발한 문법검사기는 한국어 문서에 자주 나타나는 오류를 유형별로 분류하여 의존 관계를 바탕으로 지식베이스화했다. 현재 문법 검사기에서 처리하고 있는 오류 종류 및 분류는 [6]에 설명되어 있다. 그리고 지식베이스화한 규칙을 이용해 검사 단어를 기준으로 의존문법을 이용한 부분 문장 분석으로 연어 오류 단어를 찾는 알고리즘을 개발하였다[4].

그러나 부분 문장 분석을 할 때 어휘나 의미의 중의성은 의미 오류와 문체 오류를 찾는 과정에서 다음과 같은 문제를 일으킨다.

가) 복문을 문장 분석할 때 검사단어가 속한 문장 범위를 넘어서까지 연어 오류 단어를 찾는 오류가 한 예이다.

[†] 비회원 : 부산대학교 컴퓨터공학부
kjso@solge.cs.pusan.ac.kr

^{**} 종신회원 : 부산대학교 컴퓨터공학부 교수
hckwon@hyowon.ce.pusan.ac.kr

논문접수 : 1999년 10월 18일

심사완료 : 2000년 12월 9일

나) 언어 오류 단어의 위치까지 문장 분석이 이루어지지 않고 부분 문장 분석이 끝날 수 있다.

이 두 가지 문제 때문에 전체 시스템의 교정률이 떨어지고 시스템에 대한 사용자의 신뢰가 감소한다. 어휘적 중의성은 한 단어가 두 가지 이상의 형태소 열로 분석되는 현상을 뜻한다. 이 논문에서는 부분 문장 분석 과정에서 어휘적 중의성 때문에 발생하는 문제점을 제시하고 해결 방법을 제안한다.

1.2 문서 교정 시스템에 관한 연구 현황

한국어 맞춤법 및 문법 검사기에 관한 연구는 서울대 [박종만 90] [강승식 96], 과학기술원 [강재우 90] [송춘환 90] [조영환 91] [김덕봉 90], 부산대 [채영숙 91] [심철민 96] [김현진 97] [김영진 97], 연세대 [박영환 92] [이병훈 93]을 중심으로 연구가 시작되어 현재는 여러 대학과 한글과 컴퓨터, 마이크로소프트사와 같은 소프트웨어 개발 업체를 중심으로 진행되고 있다. 또한 동아일보, 중앙일보, 연합신문 등 각 신문사에서는 문법 검사기의 중요성이 증가하여 전문가를 중심으로 대학과 연계하여 문법 검사기의 개발을 적극 추진하고 있다.

국외 문서 교정 시스템으로는 WWB, EPISTLE, CRITIQUE Ms-Word 맞춤법과 영문법 검사기 등이 있다. WWB는 Unix환경 하에서 영어 문장을 교정 처리하는 시스템으로 문서의 철자와 문법을 분석하여 사용자에게 도움말을 제공한다. EPISLE Text-Critiquing System은 영어로 쓰인 사무용 편지나 문서를 대상으로 단어, 절, 문장 단위의 오류를 처리한다. CRITIQUE는 EPISTLE시스템의 발전된 모습으로 IBM에서 개발한 시스템이다. 이 시스템은 단어, 절, 문장을 분석하여 철자 오류, 문법과 문체 오류를 찾아 주는 문서 작성 지원 도구이다. 마이크로 소프트웨어에서 개발한 Ms-Word 맞춤법 검사기는 철자 오류와 영문법 오류를 검사한다. 영문법 오류 검사 기능은 수가 일치하지 않거나 수동태 문자이나 접속사로 시작되는 문장에 대해서는 좋지 않은 문장이라는 도움말을 제공한다.

2. 문법 검사기에서 어휘 중의성에 의한 오류

철자 검사기는 현재 어절의 오류 여부를 판단하는 시스템이기 때문에 어휘적 중의성을 고려하지 않아도 된다. 그러나 문법 검사기는 검사단어를 중심으로 언어 오류 단어를 찾는 과정에서 부분 문장 분석을 해야 하므로 어휘적 중의성을 제거하지 않고는 정확한 검사가 불가능하다.

2.1 의존 문법을 이용한 부분 문장 분석

문장 분석기는 크게 어절 버퍼 관리자, 어휘 분석 모

듈, 중의성 제거 모듈, 의존 관계 결정 모듈로 구분된다. 어절 버퍼 관리자는 분석 방향에 따라 앞 어절이나 뒤 어절을 가져오고, 어휘 분석 모듈은 어절의 형태소 정보를 이용해서 의존 관계를 결정할 때 필요한 세부 정보를 저장한다. 그리고 의존 관계 결정 모듈은 의존 규칙에 따라 연속한 두 어절의 의존 관계를 조사한다. 의존 문법에서는 의존 가능한 두 가지의 어절 유형을 규정하는 것이 문법규칙이 된다[2].

본 논문에서는 의존 관계를 결정할 때 의존 관계가 성립하는 규칙보다는 의존 관계가 성립하지 않는 규칙을 이용해서 의존 관계를 결정한다. 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.

문장 분석 종료 조건(Stop-Condition)은 용언 종결형, 문장 종결 부호(!,?,.,)가 올 때다. 검사 단어와 언어 오류 단어는 대체로 같은 주어나 같은 서술어를 가지는 문장 범위 안에 존재한다. 이는 문법검사가 문장 분석 종료 조건(Stop-condition)으로 사용할 수 있는 또 하나의 기준이 된다. 검사 단어를 기준으로 문장 분석을 하던 중 '용언+연결형어미'가 나오면 문장 분석을 종료할 수 있다. 그러나 일부 용언은 동일 주어나 동일 목적어를 가진 문장의 서술어가 될 수 있다. 본 논문에서는 언어 오류 단어가 동일 주어일 때와 동일 목적어일 때 문장 분석이 실패하지 않도록 Stop-Condition에 계약을 둔다. 즉, 필수 문장 요소인 목적어나 주어 다음에 용언+연결형어미가 올 때 문장 분석을 종료한다.

다음은 언어 오류 단어가 동일 주어인 문장의 예다.

[예문 2-1] 그 형제들의 정은 두텁고 매우 깊다 (X)
(언어오류단어) (검사단어)

=> 그 형제들의 정은 두텁고 매우 깊다 (O)

위 예는 '그 형제들의 정은 두텁다'와 '그 형제들의 정은 매우 깊다' 두 문장이 접속어미 '-고'로 대등하게 연결된 문장이다. 검사단어 '깊다'부터 언어 오류 관계인 '주어'를 찾는 과정에서 문장 분석 종료 조건(stop-condition)이 될 수 있는 '두텁고'가 나오면 문법 검사기는 필수 문장 요소인 목적어 또는 주어가 이미 나왔는지 확인한다. 아직 '깊다'의 주어 가 될 수 있는 단어를 발견하지 못하였으므로 계속 문장 분석을 하여 '깊다'와 '두텁다'의 공동 주어인 '정은'을 찾는다. '정은'이

‘갑다’의 의미상 연어 오류 단어이므로 ‘갑다’를 ‘값다’로 교정한다. 동일 목적어의 처리도 같은 방법으로 한다.

다음은 연어 오류 단어가 동일 목적어인 문장의 예다.

[예문 2-2] 그는 동쪽 방향을 보고 가르쳤다 (X)
(연어오류단어) (검사단어)

=> 그는 동쪽 방향을 보고 가리켰다 (O)

서술어 ‘가르치다’와 목적어 ‘방향’은 서로 연어 오류 관계다. 검사단어 ‘가르치다’를 기준으로 앞 방향 분석을 하던 중 ‘보고’가 나오면 동일 목적어를 가질 수 있는 활용 형태인지 확인한다. 필수 문장 요소인 목적어나 서술어가 아직 발견되지 않았고 동일 목적어를 가질 수 있는 서술어이므로 계속 문장 분석을 한다. 다음 단어 ‘방향’이 ‘가르치다’의 연어 오류 단어이므로 ‘가르치다’를 ‘가리키다’로 교정한다.

그러나 의미상 공통 주어를 가질 수 없는 문장도 있다. 이 문제를 해결하는 방법은 접속어미들의 통사적 성질을 이용할 수 있다[13]. 한국어의 어떤 접속 어미들은 선행 절과 후행 절에 공통된 주어를 요구하고 어떤 접속 어미들은 선행 절과 후행 절에 독립적인 비동일 주어를 요구한다.

[예문 2-3] 그는 밥을 먹으러 식당에 간다.

[예문 2-4] 내가 증거를 제시했기에 (그 사람이) 내 말을 인정했다.

[예문 2-3]은 어미 ‘-으러’로 연결된 문장이다. ‘-으러’는 선행 절과 후행 절이 공통된 하나의 주어를 가질 때가 일반적이다. [예문 2-4]는 어미 ‘-기에’로 연결된 문장인데 후행 절의 주어가 선행 절의 주어와 같지 않다. 일반적으로 어미 ‘-기에’는 선행 절과 후행 절이 각 각 주어를 가진다.

단문장이나 주어가 동일한 병렬문은 위에서 설명한 부분 문장 분석으로 의미 오류나 문체 오류를 처리할 수 있다. 그러나 다음과 같은 문장은 부분 문장 분석으로 처리하기 어렵다.

[예문 2-5] 수업만 바뀐다고 해서 학생들에게 창의성이 손쉽게 생겨나지는 않지만, 수업이 작은 출발점이 될 수는 있다.

=> 수업이 적은 출발점이 될 수는 있다. (X)

[예문 2-6] 수업이 작은 반이 성적이 낮다

=> 수업이 적은 반이 성적이 낮다 (O)

[예문 2-5]는 “작은”이 “출발점”의 서술어인 반면 [예문 2-6]에서는 “작은”이 “수업”과 “반”을 동시에 서술하고 있다. ‘수업이 작다’는 의미상 오류가 있는 문장이다. [예문 2-5]는 ‘작다’가 ‘수업’의 서술어 역할을 하지 않으므로 의미상 오류가 없는 문장이다. 그러나 [예문 2-6

]은 ‘작다’가 ‘수업’의 서술어이므로 의미상 오류가 있는 문장이다. [예문 2-6]에서 ‘작다’를 ‘적다’로 교정해야 한다. 현재 문법 검사기가 부분 문장 분석으로 이 같은 문제를 해결하기란 어렵다. 본 논문에서는 이 같은 문장을 예외적으로 처리한다.

이 외에 부분 문장 분석이 실패하는 가장 큰 원인은 부분 문장 분석 과정에서 어휘 정보가 중의적인 단어를 처리할 때다. 의존관계를 결정할 때, 문장 분석을 종료하려고 현재 어절이 “용언+종결형”인지 결정할 때 어휘 정보가 중의적이면 그 결과를 예측하기 어렵다. 본 논문에서는 이 문제를 해결하여 문법 검사기의 검사 효율을 높일 수 있는 방법을 제안한다.

2.2 어휘 중의성 유형

자연언어를 연구하는 여러 분야에서 어휘적 중의성을 해결하려는 많은 노력이 있었다. 어휘 중의성에는 동형 이품사와 이형 동품사가 있다[14]. 문법 검사기의 부분 문장 분석 과정에서 의존 관계 설정기는 각 어절의 형태소 정보와 의존규칙을 바탕으로 입력 문장의 각 어절에 대한 의존 관계를 설정한다. 이 때 동형 이품사 때문에 구조적 중의성이 많이 발생한다. 본 논문에서는 의미·문체 검사기가 문장 분석을 수행할 때 문제가 되는 어휘 중의성을 중심으로 중의성 제거 방법을 연구한다. 다음은 문법 검사 과정에서 가장 자주 나타나는 어휘 중의성의 예를 세 가지로 분류하여 아래에 소개한다.

2.2.1 동사와 명사

동사의 어간과 어미가 결합한 형태가 명사와 형태소로 동일한 단어를 말한다.

문법검사가기 ‘에너지가 사다’는 문장을 검사하면 ‘에너지’와 ‘사다’는 주어와 용언 사이에 의미상 오류가 있다고 판단하여 ‘에너지가 싸다’로 교정한다. 그러나 ‘사다’가 활용된 형태인 ‘사실’은 동사이외에 명사 정보가 있으므로 [예문 2-8]의 문장을 검사할 때 ‘사실’을 ‘싸실’로 교정하면 안 된다.

분류1-예1) 사실

-> ① 동사: ‘사’(‘사다’의 어간)+ -시(존칭 선어말 어미)
+ -군(관형형 어미)

-> ② 명사: ‘사실’(‘실제로 있었던 일’)

해당 문장)

[예문 2-7] 어머니는 차를 사실 계획이다

[예문 2-8] 가장 큰 에너지가 임시라는 사실

‘대해’의 어휘 중의성 해결은 관용어구 규칙에 따른다. ‘대해’가 ‘대하다’의 뜻으로 사용될 때는 ‘-에 대해’로 쓰일 때가 일반적이다. 이 규칙에 따라 앞 어절의 조사가 ‘-에’이면 ‘대해’는 동사로 처리한다.

분류1-예2) 대해
 -> ① 동사 : '대하'('대하다'의 어간) + -역(연결어미)
 -> ② 명사 : '대해'('큰 바다' 뜻의 명사)
 해당 문장)
 [예문 2-9] 그는 인생에 **대해** 희망적이다.
 [예문 2-10] 그는 **대해를** 보면서 새 희망을 꿈꾼다

2.2.2 용언과 부사

동사의 활용형과 부사의 형태가 동일한 예다.

분류2-예1)보다
 -> ① 부사
 -> ② 동사 : 보('보다'의 어간) + 다(연결형어미,종결형 어미)
 해당문장)
 [예문 2-11] 그는 노력을 **보다** 많이 했다.
 [예문 2-12] 그는 TV를 **보다** 갑자기 나를 불렀다.

우리 나라의 많은 사전에서 '보다'가 부사로 쓰일 수 있다고 설명하고 있다. 그러나 우리말 큰 사전 (한글학회)에는 조사로만 쓸 수 있다고 되어 있다. '보다'를 부사로 쓰는 예는 영어 'more'와 일본어 'より'의 번역 과정에서 생긴 영향이다. 그러므로 '보다' 대신에 '더', '더욱'과 같은 우리말을 써야 바람직하다.

[예문 2-11]에서 사용된 '보다'는 부사로 사용되었으므로 '더'나 '더욱'으로 고쳐야 한다. 그러나 [예문 2-12]에서 사용된 '보다'는 [예문 2-5]에서 사용된 '보다'와는 전혀 다른 뜻인 동사로 사용되었으므로 '더'나 '더욱'으로 교정하면 안 된다.

분류2-예2) 그리고
 -> ① 부사 (접속부사)
 -> ② 동사 : 그리('그리다'의 어간) + -고(연결형 어미)
 [해당 문장)
 [예문 2-13] 그는 그림을 **그리고** 있었다.
 [예문 2-14] 음악 문학 **그리고** 무용은 모두 예술의 한 분야이다.

위 예에서 나타난 '그리고'는 '그리다'의 활용형과 접속부사 두 가지로 해석할 수 있다. '그리고'가 접속부사인지 알 수 있는 방법은 문장의 병렬성을 이용해야 한다. 접속 구문의 분석은 자연어의 분석에 있어서 매우 어려운 문제의 하나로, 접속구문의 구조적 중의성 문제를 해결하려는 연구가 대등 접속 구문의 대등성 혹은 병렬성에 맞추어 이루어졌다[12]. 그러나 한 문장 속의 연속한 임의의 두 단어열이 서로 비슷함을 기계적으로 발견하기란 쉽지 않다. 접속 구문의 대등성과 병렬성을 분석하려면 전체 문장의 분석이 이루어져야 하므로 본 시스템에서 접속부사와 동사의 어휘적 중의성을 완벽하게 해결하기는 어렵다. 다만 '그리고'에 대해서 앞 어절

에 목적어나 주어를 취하면 '동사'로 보는 방법으로 어휘적 중의성을 최소화하고 있다.

분류2-예3)매달
 -> ① 부사 :
 -> ② 용언: 매다('매달다'의 어간)+ 리(관형형어미)
 [해당 문장)
 [예문 2-15] 수출이 **매달** 증가하였다.
 [예문 2-16] 그는 돈에 목 **매달** 만큼 가난하지 않다.

부사와 용언의 어휘 중의성은 문법 검사기가 올바른 문장을 비문으로 간주하여 문장 분석을 제대로 하지 못하는 요인이 되고 있다. [예문 2-15]에 사용된 '매달'은 부사로 사용되어 '증가하였다'를 수식하고 있다. 문법 검사기가 [예문 2-15]의 문장에서 '증가하였다'의 주어를 찾으려고 문장 분석을 시도할 때 '매달'을 '동사+관형형어미'로 해석할 수 있다. 의존 관계를 이용한 문장 분석 과정에서 지배소가 "용언(증가하다)"이고 의존소가 "동사+관형형어미(매달)"이면 의존 관계가 성립하지 않으므로 이 문장을 비문으로 간주하는 오류가 생긴다.

2.2.3 동사와 조동사

예를 들면, '하다'는 본동사로도 쓰이지만 앞에 '-어야'나 '-아야' 같은 어미가 오면 보조 용언의 기능을 한다.

이 때, 보조용언 '-어야/아야 하다'라는 패턴정보를 규칙에 넣지 않으면, '하다'의 의존소는 '정책'으로 간주한다. 그런데 '정책을 하다'와 '정책을 펴다'가 오류-표준어 쌍으로 검사대상에 등록되어 있기 때문에 교정기는 '하다'를 '펴다'로 잘못 교정할 수 있다. 이와 같이 '-어야/아야 한다', '-고/을 수 있다'와 같이 동시에 본 용언과 보조용언으로 쓰일 수 있는 동사는 표현 자체를 하나의 관용어구로 묶어서 처리한다.

분류3-예1) 하다
 -> ① 보조용언 (어미 '-어야/아야'를 동반)
 -> ② 본용언
 해당 문장)
 [예문 2-17] 새로운 정책을 과거청산 방법으로 보아야 **한다**.

2.3 문법 오류를 처리할 때 어휘적 중의성에 의해 발생하는 문제점

어휘 중의성이 있는 단어를 처리할 때 발생하는 문제는 문법 검사기의 모듈에 따라 다르다. 검사 단어의 어휘가 중의적일 때 발생하는 '규칙 전처리 오류'와 문장 분석 과정에서 분석 중인 단어의 어휘가 중의적일 때 올바른 문장을 비문으로 간주하거나 단문을 이중 문장으로 간주하여 언어 오류 단어를 찾기 전에 문장 분

석을 종료하는 '문장 분석 오류'가 있다.

다음은 '문장 분석 오류' 예다.

문장 분석 오류 예	
규칙 1 (품사 혼동 오류) 검사 단어 : 보다 언어 오류 단어 : 명사 및 용언 대치 규칙 : 앞 명사와 붙이기	규칙 2:(문체 오류) 검사 단어 : 보다 언어 오류 단어 : 부사 대치 단어 : 더욱
[예문 2-18] 그는 인생에 대해 보다 긍정적이다 <각 어휘 형태소 분석 결과> [('보다(부사)') [('보(명사)' + '다(조사)')] [('보(동사)' + '다(종결어미)')] [('대해(명사)') [('대하(동사)' + '여(연결어미)')] [('인생(명사)' + '에(조사)')] <의미 문체 오류 처리 과정>	
<div style="border: 1px solid black; padding: 5px; margin: 10px auto; width: fit-content;"> <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px; display: inline-block;"> 검사 단어 : 보다 언어 오류 단어 : 명사 </div> <div style="display: flex; justify-content: space-around; align-items: center;"> 그는 인생에 <div style="border: 1px solid black; padding: 2px;"> 대해 보다 </div> 긍정적이다 </div> <p style="text-align: center; margin-top: 5px;">규칙 1 적용 => 그는 인생에 대해보다 긍정적이다</p> </div>	

위 예문은 문체 오류가 있는 문장이다. '보다'를 부사로 쓰는 예는 영어 'more'와 일본어 'より'의 번역 과정에서 생긴 영향이다. '보다' 대신에 '더', '더욱'과 같은 우리말을 써야 바람직하다(규칙2 적용). 그러나 위 예문의 오류를 검사하는 과정에서 앞 어절이 명사('대해')이므로 문법 검사기는 조사인 '보다'와 명사 '대해' 사이에 띄어쓰기 오류가 있다고 검사하여(규칙1 적용) '대해보다'로 잘못 교정한다. 이는 '대해'의 어휘 중의성 때문에 발생한 문법검사 검증 오류이다.

문법 검사기는 철자 검사와 의미·문체 검사를 순서적으로 수행한다. 철자 검사 과정에서 현재 어절이 의미·문체 검사 오류 가능성이 있는 어절이면 나중에 의미·문체 검사를 할 수 있도록 어절의 어간과 형태소 정보를 저장한다. 어휘 중의성이 있는 검사 단어는 여러 개의 표제어와 어휘 정보를 가질 수 있다. 철자검사는 여러 개의 표제어 중 가장 짧은 표제어와 이 표제어를 어간으로 한 형태소 정보를 의미 문체 검사기로 넘긴다. 의미 문체 검사기의 규칙 검색부는 어절의 어간(Root)을 표제어로 규칙베이스에서 규칙을 찾는다. 그러므로 어간의 길이가 다르게 분석되는 어절은 규칙 처리가 힘들다. 이를 해결하려면 서로 다르게 분석되는 모든 어간

에 대해서 규칙을 적용해야 한다. 규칙에는 의미나 문법 오류 유형에 따라 검사단어의 품사가 명시되어 있다. 문법 검사기의 규칙 검사부는 규칙에 명시한 검사단어의 품사와 현재 검사단어의 품사가 같은지 검증한다. 이때 현재 검사단어의 어휘 중의성을 제거하지 않으면 규칙 검사부의 검사 효율이 떨어진다. 이를 '규칙 전처리 오류'로 구분한다.

다음은 규칙 전처리 오류의 예다.

규칙 전처리 오류 예
규칙 233 검사 단어 : 하다 검사 단어 형태소 정보 : 동사 언어 오류 단어 : 정책 대치 단어 : 피다
[예문 2-19] 새로운 수상은 수출 지향 정책을 했다.
[예문 2-20] 새로운 정책을 과거청산 방법으로 보아야 한다.

[예문 2-19]는 검사단어 '하다'가 본용언으로 사용되었으므로 '정책'과 '하다' 사이에 의미 오류가 발생했다. [예문 2-20]은 '하다'가 보조용언으로 사용되어 '정책'의 직접 서술어가 아니므로 '정책'과 '하다' 사이에 의미 오류가 없는 문장이다. 이 두 문장을 처리할 때 '하다'의 어휘 중의성을 해소하지 않으면 문법 검사기가 [예문 2-20]을 "새로운 정책을 과거청산 방법으로 보아야 한다"로 잘못 교정하게 된다.

3. 문법 검사기에서 중의성 해결 방법

현재 개발된 대부분의 철자 검사기에는 품사 태깅 시스템이 부착되어 있지 않다. 본 논문에서는 품사 태깅 시스템 없이 의미 오류나 문체 오류를 검사하는 과정에서 경험적 규칙을 이용하여 어휘 중의성을 제거하고자 한다. 본 논문에서 사용하는 경험적 규칙은 좌·우 어절의 어휘 정보를 이용하여 현재 어절의 형태소 정보를 구한다. 이는 한국어에서 자주 나타나는 어절 간 수식관계를 분석하여 구한 정보다.

문법 검사기에서 어휘 중의성을 해결하려는 목적은 문법 검사 처리 성능의 향상이다. 한국어에서 어휘 중의성을 제거하는 방법은 확률을 이용하는 방법, 퍼지망을 이용하는 방법, 신경망을 이용하는 방법 등 다양하다. 이 같은 중의성 제거 방법을 사용할 때 어휘 중의성 제거율은 높아지지만 문법 검사기에 적용하기는 구현이 어렵고 문법 검사기의 검사 속도를 떨어뜨릴 수 있다.

3.1 중의성 제거 규칙 사용

문장 분석과정에서 형태소적 중의성이 있는 단어를 만나면 어절 간 의존관계를 결정할 때 예측하기 어려운 결과가 나올 수 있다. 본 논문에서는 문장 분석 과정에서 중의성 제거 규칙을 이용해 중의성을 제거한다. 한국어에서 인접한 두 어절을 살펴볼 때, 앞 어절의 어말 부분이 뒤 어절과 문맥적 관련성이 깊고 뒤 어절의 어두 부분이 앞 어절과 문맥적 관련성이 깊다. 한국어의 이런 특성 때문에 어말-어두 공기 정보를 이용한 문맥 확률을 이용하여 어휘 중의성을 해소할 수도 있다[8]. 본 논문에서는 문법 검사기가 의미나 문체 오류를 처리할 때 자주 문제가 되는 어휘 중의성 유형을 중심으로, 이를 해결할 수 있는 문맥 정보를 경험적 언어 지식에 의해 만들었다. 이를 “중의성 제거 규칙”이라 한다.

본 논문에서 사용하고 있는 규칙의 유형은 어휘별 중의성 제거 규칙과 일반적 중의성 제거 규칙으로 분류할 수 있다. 어휘별 중의성 제거 규칙은 입력 문장의 각각의 어휘에 대해 규칙이 존재한다. 예를 들면 앞 어절의 어미가 ‘-어아’, ‘-아아’, ‘-해야’로 끝난 용언일 때 ‘하다’는 보조용언이다. 그리고 일반적 중의성 제거 규칙은 모든 입력 문장에 대해 동일하게 적용되는 규칙으로 먼저 분석된 왼쪽 또는 오른쪽 어절의 품사 정보를 이용해서 현재 어절의 품사 정보를 결정한다. 예를 들면 왼쪽 문장 분석을 할 때 먼저 분석된 어절이 용언이고 왼쪽에 이웃한 어절이 동사+관형형, 명사+목적어로 중의적 어휘정보를 가질 때 ‘동사+관형형’보다는 ‘명사+목적어’를 우선 선택한다. 규칙을 적용하는 과정은 중의성이 있는 어절을 중심으로 좌우 문맥 정보를 보고 중의성을 제거한다.

본 논문에서 제안하는 규칙은 어두와 어말의 문맥 연관성을 이용하므로 어두와 어말의 범주를 결정해야한다. 본 시스템이 문장 분석 과정에서 사용하고 있는 어휘 정보는 표 1과 같다.

한국어에서 어두는 매우 다양하므로 너무 세분화하면 규칙의 유형이 실제 문장에서 나타나지 않을 수도 있다. 따라서 본 논문에서는 중의성 제거 규칙을 설정할 때 사용하는 어두의 범주를 명사, 본용언, 보조용언, 술어부사, 부사, 관형사, 수사 7가지로 제한한다. 그러나 어말은 그렇게 다양하지 않으므로 범주를 구체적으로 세분화해서 사용해도 규칙의 유형이 문장에 나타나지 않을 가능성이 적다[8].

가) 어휘별 중의성 제거 규칙

(Specific Disambiguation Rule)

어휘별 중의성 제거 규칙이란 문장에서 나타나는 각각의 어휘에 대해서 중의성 제거 방법이 존재하는 규칙

표 1 문법 검사기에서 사용하는 세부 형태소 정보

품사	세부 형태소 정보	예
명사	명사 + 주격 조사	학생은, 학교를
	명사 + 목적격조사	학생도, 학생은
	명사 + 을/를	학교를
	명사 + 소유격조사	학생의
	명사 + 접속격조사	학생과
	명사 + 부사격조사	학생에게
	명사 + '이다' 축약 조사	공부다
	하다 결합명사 되다 결합명사 이다 결합명사	공부하다 치리되다 공부이다
본용언	동사 + 관형형어미	먹는
	동사 + 연결형어미	먹고
	동사 + 부사격전성어미	먹게
	동사 + 종결형어미	먹다
	형용사 + 관형형어미	예쁜
형용사 + 연결형어미	예쁘고	
형용사 + 부사격전성어미	예쁘게	
형용사 + 종결형어미	예쁘다	
술어부사	술어부사	불과하다
	술어부사 + 관형형어미	불과한
	술어부사 + 연결형어미	불과하고
	술어부사 + 부사형어미	불과하게
	술어부사 + 종결형어미	불과하다
부사	부사	매우
관형사	관형사	몇몇
수사	숫자로 시작되는 수사	2백
기타	용언+ 어미	들러, 들리는데

이다. 어휘별 중의성 제거 규칙은 다음과 같은 형태로 구성된다.

RULE(#)=(WORD, Direct, Condition, Selection)

어휘별 중의성 제거 규칙은 규칙에 명시된 특정 어휘가(WORD) 나타날 때만 적용된다. 중의성이 있는 특정 어휘(WORD)가 입력되면 이웃한 왼쪽어절이나 오른쪽 어절이 규칙에서 제시한 조건(Condition)을 만족하는지 검증한다. 조건을 만족하면 현재 단어는 Selection에 명시한 품사로 해석한다.

예로 용언 ‘하다’가 본용언과 보조 용언으로 해석될 때 중의성을 제거하는 규칙과 ‘되도록’이 동사+어미와 부사로 해석될 때 중의성을 제거하는 규칙은 다음과 같다.

RULE(1)=(하다, LEFT, e 어아 | e 아아 | e 야, PX)
 RULE(2)=(되도록,LEFT, j 이 | j 가, PV)
 RULE(3)=(되도록,RIGHT, MM|MA, MA)
 LEFT : 왼쪽 어절
 RIGHT : 오른쪽 어절
 (E :어미, J :조사, MM :관형사, MA:부사
 PX:보조용언, PV: 본용언)

본용언 '하다'는 정책과 연어 오류 관계다. '정책을 하다'보다는 '정책을 펴다'로 쓰는 것이 바람직하다. 그러나 [예문 2-17]에서 사용된 '하다'는 왼쪽 어절의 어머니가 '-이야'가 나왔기 때문에 보조 용언으로 해석되어 '정책'과 연어 오류 단어가 될 수 없다. 그러나 [예문 3-1]에서는 '하다'가 본용언이고 연어오류단어 '정책을'이 발견되었으므로 '하느'을 '펴느'으로 교정한다.

[예문 3-1] : 인간중심 정책을 하는 정치인 (x)
(연어오류단어) (검사단어)

=>인간중심 정책을 펴는 정치인 (o)

나) 일반적 중의성 제거 규칙

(General Disambiguation Rule)

이 규칙은 문장 분석을 할 때 이미 분석된 어절의 품사를 이용해서 그 다음 입력된 어절의 품사를 결정하는 방법이다. 즉 검사 단어를 시작으로 문장 분석 방향에 따라 입력된 어절의 어휘가 중의적일 때 앞에 분석된 어절의 품사를 이용해서 중의성을 제거하려는 방법이다. 문법 검사기는 검사단어의 연어오류를 찾는 과정에서 앞 또는 뒤 방향으로 연속하는 어절의 형태소 정보를 저장하고 있다. 즉, 이미 분석된 정보를 이용해 현재 단어의 중의성을 제거하기 때문에 문장 분석 과정에서 중의성 제거에 필요한 시간이 많이 늘어나진 않는다.

일반적 중의성 제거 규칙은 다음과 같은 형태로 구성된다.

RULE(#)=

(중의성유형,

: (Left,Condition,Col/Anti),(Right,Condition,Col/Anti),
Selection,)

'중의성 유형'은 중의적으로 분석된 형태소 정보를 나타내고 condition은 중의성을 제거하는데 필요한 왼쪽 (Left) 또는 오른쪽(Right) 어절의 품사 정보를 나타낸다. Col/Anti는 규칙과 주어진 조건과의 관계를 나타낸다. Anti_collocation(Anti)은 왼쪽이나 오른쪽 어절이 Condition에 명시된 품사가 아닐 때 조건을 만족하고, Collocation(Col)은 왼쪽이나 오른쪽 어절이 Condition에 명시된 품사로 분석될 때 조건을 만족한다, 조건이 만족하면 규칙에 의해 Selection에 명시된 품사를 선택한다.

예로 명사+ 조사와 용언+관형형어미로 해석되는 어절에서 중의성을 제거하는 규칙은 다음과 같다.

RULE(4)

(N+J & P+etm,(Left,0,0),(Right, N, Anti), (N+J)N);

RULE(5)

(N+J &

P+etm,(Left,(P+etm)N|(N+jcm),Anti),(Right,N|N+j,
Col),P+etm);

LEFT : 왼쪽 어절, RIGHT : 오른쪽 어절

N+J :명사+조사, N+jcm : 명사+소유격조사

N:조사가 생략된 명사

P+etm :용언 + 관형사형어미

현재 어절 오른쪽에 있는 어절이 명사가 아니면 현재 어절을 명사+조사나 조사가 생략된 명사로 본다. 오른쪽 어절이 명사고, 왼쪽 어절이 명사+소유격조사나 명사 또는 용언+관형형어미가 아닐 때 현재 어절을 용언+관형형어미로 본다.

검사단어 : 부치다

연어 오류 단어 : 불, 감시원, 꼬리표, 마음, 말 등

대치단어 : 붙이다

[예문 3-2] 어머니가 화로에 불을 부치고 계셨다.

(N+jco); (P+etm)

=>어머니가 화로에 불을 붙이고 계셨다

[예문 3-2]에서 '붙을'은 명사+목적격조사와 동사+관형형어미로 해석된다. 그러나 '부치다'가 용언이기 때문에 RULE(4)에 의해서 명사+목적격조사로 해석되고 '부치다'와 연어 오류 단어가 된다. 문법 검사기는 연어 오류 관계 단어가 발견되었기 때문에 '부치다'를 '붙이다'로 교정한다.

3.2 어휘 중의성 제거 규칙을 이용한 부분 문장 분석

문법 검사기의 규칙 검증 처리부는 의미·문체 오류를 검사하려고 검사 단어를 기준으로 왼쪽, 오른쪽에 연어 오류 관계가 있는 단어를 조사한다. 검사단어와 연어 오류 단어 사이에 여러 어절이 있을 때는 문장 분석기에서 의존 문법에 기반한 부분 문장 분석을 한다.

다음은 어휘 분석결과를 이용하여 의존 문법으로 검사단어의 왼쪽 또는 오른쪽 방향으로 문장 분석을 하는 알고리즘이다.

(3)에서 어휘 분석을 하여 (4)에서는 형태소적인 중의성이 있으면 경험적 규칙을 이용해서 중의성을 제거한다. 처리 단계 (5)~(10)에서 입력 어절의 지배소를 찾아 (11)에서 의존 관계를 분석하여 결과를 저장한다. 새로운 어절의 지배소를 찾을 때는 문장의 투영성[2]을 보장한다. 이미 의존 관계가 설정된 어절은 지배소의 위치를 확인한다. 만약 지배소가 존재하지 않으면 문장 분석을 끝낸다.

문법 검사기는 하나의 표제어에 대해 여러 개의 규칙을 적용할 수 있다. 규칙에 따라 연관 관계가 있는 어절의 위치가 다르므로 문장에서 분석할 어절 수는 달라질

```

1) 검사 단어로부터 왼쪽 또는 오른쪽 방향의 i번째
   어절(WORD(I))을 가져 온다
2) if (의존 관계가 설정 안 된 어절이면) {
3) MORP(I) =CHECK-ALL-INFO(WORD( I ))
4) RMAMBIGUITY(MORP(I))
5) for (IndexOfGov=i-1; IndexOfGov >=0; ) {
6)   If (MORP(IndexOfGov)와 MORP(I) 의존관계 성
   립){
7)     GOV(I)= IndexOfGov
8)     break;
9)   }else
10)    IndexOfGov = GOV(IndexOfGov)
11)   IndexOfGov=NOT_EXIST
12) } // end for
13) STOREArray(WORD(I), MORP(I), Gov(I))
14) }else
15)   IndexOfGov= RETRIVEArray(I)
16) if( IndexOfGov != NOT_EXIST) return TRUE
17) else return STOP_NOW
    
```

그림 1 부분 문장 분석 알고리즘

수 있다. 그러나 이미 의존 관계가 설정된 어절은 다시 분석하지 않는다.

다음은 의존 문법을 이용한 부분 문장 분석 과정 및 오류 교정의 예를 보인다.

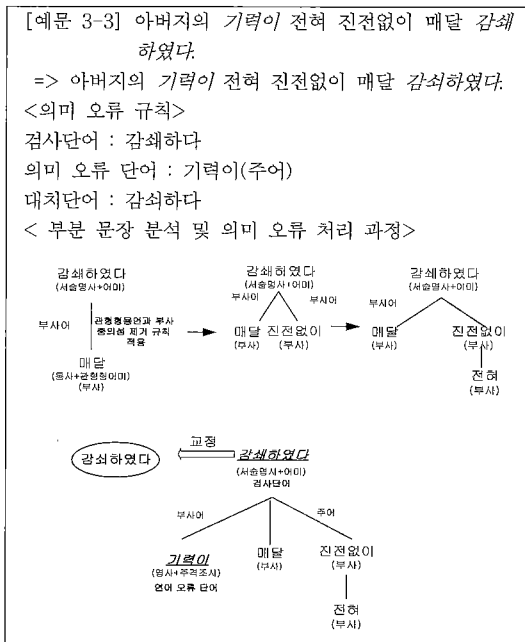


그림 2 어휘 중의성을 고려한 부분 문장 분석 과정

'감쇄하다'는 '덜러 없어지거나 떨어져 없애다'는 의미이고, '감쇄하다'는 '힘이나 세력 따위가 줄어서 약해지다'는 뜻이다. 즉 '감쇄하다'는 의미상 '기력'의 서술어가 될 수 없다. 문법 검사기는 의미 오류를 유발할 수 있는 검사 단어 '감쇄하다'를 기준으로 문장에 오류가 있는지 검사한다. 의미 오류 처리 규칙에 '감쇄하다'와 연어 오류 관계가 있는 단어의 문장 성분이 주어이므로 앞 방향으로 문장 분석을 한다. 문장 분석을 하던 중 주어 '기력'이 나오면 어간 '기력'이 검사단어의 연어 오류 단어인지 검사하여 의미 오류가 발생하면 교정한다.

4. 문법 검사기 시스템 구성도

문법 검사기에서 의미 오류와 문체 오류를 검증하는 부분은 규칙 검색부, 규칙 전처리기, 규칙 처리부, 오류 단어 교정부로 구성되어 있다[6]. 규칙 처리부에서는 연어 오류 단어와 검사 단어 사이에 여러 어절이 있을 때 부분 문장 분석 모듈을 호출한다. 어휘 중의성 제거규칙은 규칙 전처리부와 부분 문장 분석 모듈에서 이용된다.

4.1 의미 오류 처리부의 중심 모듈

(1) 지식베이스 검색부

검사 단어에 적용될 N개의 의미 오류 처리 규칙을 지식베이스에서 찾아 그 시작 위치를 리턴한다.

(2) 규칙 전처리기

규칙을 검사하고 교정하는 과정에서 편리하도록 의미 오류 처리 규칙 내용을 구조체 변수에 대입한다. 규칙에서 제안하는 검사 단어의 어휘 정보와 현재 처리 중인 검사 단어의 어휘 정보가 일치하는지 검증한다. 이 때 검사 단어의 어휘가 중의적이면 중의성 제거 규칙으로 검사 단어의 어휘 중의성을 제거한다.

(3) 규칙 처리부

규칙 전처리기에서 세팅한 구조체 값에 따라 연어 관계 단어가 있는지 검색하고 있으면 그 단어가 연어 오류 관계 단어인지 본다. 규칙 처리부는 연어 오류 단어의 문장 성분에 따라 크게 6개의 검증부로 구분된다. [그림 3]은 규칙 처리부의 세부 모듈을 나타낸다.

(4) 오류 단어 교정부

의미 오류가 발생하면 오류 어절(검사 단어)을 대치 단어로 교정한다. 대치 단어와 교정 규칙 정보는 규칙 전처리기에서 설정한 값을 사용한다.

4.2 구분 분석을 위한 모듈 상호 관계

검사 단어와 연어 관계 단어 사이의 문장을 분석할 때 필요한 모듈 및 상호 관계는 [그림 4]에 나타나 있다.

의미·문체 규칙 처리부는 의미·문체 지식베이스에

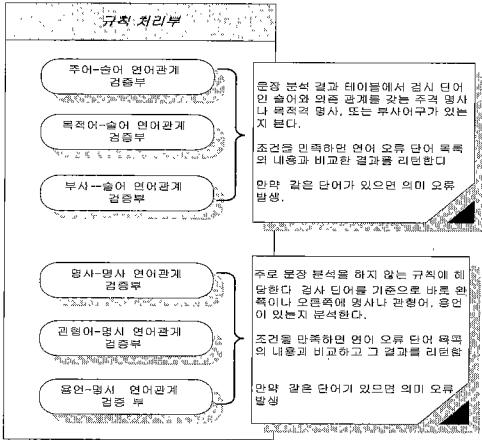


그림 3 규칙 처리부

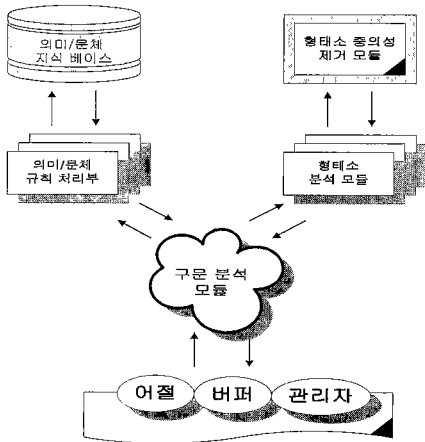


그림 4 부분 문장 분석을 위한 모듈간 상호 관계

서 검사 단어에 해당하는 규칙을 검색한다. 규칙에 따라 부분 문장 분석 모듈을 호출한다.

부분 문장 분석 모듈은 어절 버퍼 관리자로부터 현재 기준 단어를 중심으로 왼쪽이나 오른쪽으로 N번째 위치한 어절을 제공받는다. 의존 관계를 설정하려면 N번째 어절의 어휘 정보가 필요하므로 형태소 분석 모듈을 호출한다.

형태소 분석 모듈은 어절의 전자 사전 정보를 중심으로 실질 형태소와 형식 형태소의 품사를 규정한다. 이때 '이다'명사, '하다'명사, '되다'명사 등이 뒤에 '이다', '하다', '되다'와 결합할 때에는 용언으로 분석한다. 만약 N번째 어절의 어휘가 중의적이면 중의성 제거 모듈에서 중의성 제거 규칙을 이용하여 N번째 어절의 어휘 중의

성을 제거한다. 형태소 분석 모듈에서 구한 어휘 정보와 의존 문법을 기반으로 지배소와 의존소 간의 의존 관계를 설정한다.

의미·문체 규칙 처리부는 부분 문장 분석 결과와 지식베이스에 기반해서 후보 언어 관계 단어(즉 언어 관계 단어와 같은 품사 정보를 가지는 단어)가 언어 오류 단어인지 비교한다. 의미 오류 또는 문체 오류가 발생하면 지식베이스의 교정 규칙에 따라 교정을 수행한다.

5. 실험

이 장에서는 앞에서 제안한 의존 문법을 이용한 부분 문장 분석과 중의성 제거 규칙을 이용한 문법 검사기의 효율성을 보인다. 그러나 원래 문서에서 의미 오류와 문체 오류가 얼마나 발생하는지 정확하게 알 수 없으므로 재현율과 정확도를 측정할 수는 없다.

5.1 실험 대상 자료 및 환경

이 논문에서는 교과서처럼 이미 검열된 문서보다는 개인적으로 주고 받는 문서를 이용해서 의미 오류와 문체 오류의 유형을 살펴보고 문법 검사기가 제시하는 대치어가 문맥상 맞는지 검증한다. 실험에 사용된 자료는 모두 pc통신 자료로서 문서의 성격에 따라 분류하여 실험하였다. 실험을 위해 수집된 데이터의 종류는 다음과 같다.

분야	문장 수	어절 수	용량
소설	11,734	76,141	602 kb
논문	2,505	34,325	242 kb

그리고 의미 및 문체 지식베이스에 있는 규칙의 수는 총 2,300개다. 구문 분석과정에서 사용할 중의성 제거 규칙 수는 어휘별 중의성 제거 규칙 8개와 일반적 중의성 제거 규칙 9개가 사용되었다.

5.2 실험 결과

문법 검사기의 평가 기준으로는 검사 성공률을 측정한다. 측정을 위한 합수는 다음과 같다. 검사 성공률은 전체 오류 어절 중 오류 판정이 맞는 어절수의 비율로 정의된다.

$$\begin{aligned}
 &C_{total} : \text{전체 오류 어절 수} \\
 &C_{fail} : \text{오류 유형 추정 어절 수} \\
 &\text{검사 성공률 :} \\
 &P(c) = \frac{(C_{total} - C_{fail})}{C_{total}} \times 100
 \end{aligned}$$

[표 2]는 실험을 통해 나타난 검사 결과를 나타내고 있다. 여기서 A system은 구문 분석 과정에서 중의성 제거규칙을 사용하지 않은 시스템이고 B system은 중

의성 제거 규칙을 이용해서 문법 검사를 수행한 시스템이다.

표 2 문법 검사기가 발견한 오류 어절 수

분야	시스템 mode	검사 성공	검사 실패	검사 성공률
소설	A system	239	30	89%
	B system	259	10	96%
논문	A system	139	14	91%
	B system	151	2	99%

[표 2]에서 소설보다 논문에서 검사 성공률이 높게 나타났다. 그 원인을 살펴보면 논문에서 사용하는 어절은 대부분 정제된 어절이고 오류의 유형이 단순하거나 일정하므로 오류 검사 성공률이 높다. 반면에 소설은 미등록어나 오류 유형이 불확실한 어절이 많이 나타나므로 검사 성공률이 낮다.

실험 결과에서 나타난 검사 실패 요인으로는 연어 오류 가능 단어의 형태소적 중의성 때문에 생긴 검사 실패와 규칙 제약 조건 오류 및 연어 오류 단어의 의미적 중의성이 주요 요인이다. A system과 B system은 다른 모든 환경은 같고 중의성 제거 규칙을 이용한 구문 분석 적용 여부만 다르게 하였다. A system은 구문 분석 과정에서 형태소적 중의성이 있는 어절 때문에 오류를 찾지 못한 오류 수가 20개 발견되었다. 그리고 B system에서는 형태소적 중의성을 제거한 후 구문분석한 결과 검사 성공률이 A system보다 소설에서는 7% 향상되었고 논문에서는 8% 향상되었다.

6. 결 론

어떤 문장은 몇 번을 읽어도 내용을 이해할 수 없을 때가 많다. 이런 문장을 가능한 한 문법과 오류가 없으며 이해하기 쉬운 문장으로 교정해 나가는 것이 의미 / 문체 오류 처리의 기본 목적이다. 나아가 각 오류에 대한 상세한 도움말과 교정 예를 제시하여 문법 검사기의 사용자가 우리말과 글에 대한 정확한 사용법과 의미를 학습하고 그 중요성도 인식할 수 있도록 한다. 그러나 한 어절 단위의 철자 오류나 띄어쓰기 오류를 처리하는 철자 검사·교정기의 기능과는 달리 여러 어절을 이용해서 오류 분석을 하는 문법 검사기는 문장 분석을 해야한다. 이 논문에서는 의존 문법을 이용한 부분 문장 분석을 통해 연어 오류 단어를 찾는다. 문장 분석에서 지배소와 의존소의 의존 관계를 결정하려면 어절의 세부적 형태소 정보가 필요하다. 그러나 한 어절은 여러

개의 형태소 정보를 가질 수 있다. 이 논문에서는 형태소적 중의성이 있는 단어를 이용해 의존 관계를 결정할 때 중의성 제거 규칙을 사용하였다.

의미 오류가 전체 문서에서 나타나는 횟수는 철자 오류나 표준어 오류에 비해서 적다. 그러나 의미 오류나 문체 오류는 사용자의 실수로 생긴 오류가 아니고 사용자가 몰라서 발생할 가능성이 큰 오류이므로 문서에 나타난 횟수는 적어도 의미 오류나 문체 오류를 검사 교정하는 일은 매우 중요한 작업이다.

향후과제로는 한 검사 단어의 두 가지 규칙에서 동시에 연어 오류 단어를 찾았을 때 발생하는 규칙 충돌 문제와 부분 문장 분석으로 처리하기 어려운 내포문에서 연어 오류 단어의 검사 방법이 해결되어야 한다. 그리고 앞, 뒤 어절의 품사를 이용한 중의성 제거 규칙은 앞, 뒤 어절의 어휘적 중의성이 제거되지 않았을 때는 적용할 수 없다는 단점이 있으므로 이 문제가 해결되어야 한다. 그리고 검사 과정에서 연어 오류 관계일 가능성이 있는 어절이 여러 개의 의미를 가질 때 어떤 의미로 사용되는지 알 수 있어야 한다.

참 고 문 헌

- [1] 박용옥, 조혁규, 권혁철, 의존 문법을 이용한 한국어 분 서기의 구현, 90정보과학회 봄 학술발표논문집, pp. 191-194, 1990
- [2] 손광주, 홍영국, 이종혁, 이근배, 어절간 의존관계 해석을 위한 한국어 파서, HCI 94 발표논문집, pp. 135-136
- [3] 홍영국, 이종혁, 한국어 의존 해석을 위한 형태소-통사적 품사 분류 체계, 정보과학회논문지(B) 제22권 제9호, pp 1375-1383
- [4] 김영진, 최성필, 손훈석, 박용옥, 권혁철, 단어의 하위 범주화 정보를 이용한 한국어 문법 검사기, 97 인공지능연구회 춘계 학술발표논문집, pp 72-75
- [5] 최명석, 이공주, 김길창, 영한 기계 번역 시스템에서 휴리스틱을 이용한 영어 문장의 구조적 중의성 해소, 97 한국정보과학회 가을 학술발표논문집, pp. 178-180
- [6] 김현진, 어절 간 연관 관계를 이용한 한국어 문법 검사기, 정보과학회논문지,
- [7] 권혁철, 윤애선, 최준영, 단일화 기반 의존 문법에 의한 자연언어 분석 기법, 한국정보과학회 봄 학술발표 논문집, 1991
- [8] 이하규, 어말-어두 공기 정보를 이용한 한국어 어휘 중의성 해소, 한국정보과학회논문지(B) 제24권 제1호, pp. 82-89
- [9] 채영숙, 연어 규칙에 기반한 한국어 문서 교정 시스템의 구현, 부산대학교 박사학위 논문, 1998
- [10] 김민정, 규칙과 말뭉치를 이용한 한국어 형태소·분석과 중의성 제거, 부산대학교 박사학위 논문, 1997

- [11] 양재형, 다중 지식원을 이용한 한국어의 분석, 한국정보과학회논문지, 제21권 제7호, pp. 1324-1332
- [12] 윤준태, 한국어의 대등접속구문 분석, 한국정보과학회 논문지, 제24권 제3호, pp.326-335
- [13] 이은경, 국어의 접속 어미 연구, 국어연구 제97호, 국어연구회, 1990
- [14] 이상주, 은닉 마르코프 모델을 이용한 두 단계 한국어 품사 태깅, 1994
- [15] 강승식, 장병탁, 음절 특성을 이용한 범용 한국어 형태소 분석기 및 맞춤법 검사기, 정보과학회 논문지, 제 23권 제 5호, 1996
- [16] 강제우, 접속 정보를 이용한 한글 철자 및 띄어쓰기 검사기의 설계 및 구현, 한국과학기술원 석사학위논문, 1990
- [17] 송춘환, 원형 복원 방법을 이용한 한글 철자 검사기의 설계 및 구현, 한국과학기술원 석사학위논문, 1990
- [18] 조영환, 한글 맞춤법 교정기의 설계 및 구현, 한국과학기술원 석사학위논문, 1991
- [19] 김덕봉, 최기선, 강제우, 한국어 형태소와 사전-접속 정보를 이용한 한글 철자 및 띄어쓰기 검사기, 언어연구, 제26권 제1호, pp87-113, 1990
- [20] 채영숙, 김재원, 김민정, 권혁철, 한국어 철자 검색을 위한 형태소 분석 기법, 91 우리말 정보화 잔치 국어 정보학회, pp179-186, 1991
- [21] 심철민, 권혁철, 연어 정보에 기반한 한국어 철자 검사기와 교정기의 구현, 정보과학회 논문지, 제23권 제7호 1996
- [22] 박영환, 말뭉치에 기반한 형태소 분석기 및 철자 검사기의 구현, 연세대학교 석사학위 논문, 1992
- [23] 이병훈, 윤준태, 송만석, 말뭉치를 기반으로 한 한국어 철자 교정기의 구현, 한글 및 한국어 정보처리 학술발표논문집, 1993
- [24] Allen James, 1995, Natural Language Understanding, Second Edition, The Benjamin/Cummings Publishing Company, INC.
- [25] Hyuck Christian R. & Lytinen Steven L., 1993, Efficient Heuristic Natural Language Parsing, Proceedings of the 11th National Conference on Artificial Intelligence, July 11-15, 93, Washington, DC.
- [26] Keller Frank & Zechner Klaus, 1995, A connectionist Model of Lexical and Contextual Influences on Ambiguity Resolution in Human Sentence Processing, Proceedings NLP95, December 4-7, 95, Volume 2, Seoul.[1] 강승식, 장병탁, 음절 특성을 이용한 범용 한국어 형태소 분석기 및 맞춤법 검사기, 정보과학회 논문지, 제 23권 제 5호, 1996



소길자

1994년 동의대학교 전산통계학과 학사.
1999년 부산대학교 전자계산학과 석사.
1999년 ~ 현재 부산대학교 시간강사. 관심분야는 자연언어처리



권혁철

1982년 서울대학교 공과대학 전산학 학사. 1984년 서울대학교 공과대학 전산학 석사학위 취득. 1987년 서울대학교 공과대학 전산학 박사학위 취득. 1988년 ~ 부산대학교 정보컴퓨터 공학부 교수 1992년 ~ 한국 정보과학회 운영위원. 1992년 ~ 1993년 미국 standfordeogkr CSLI연구소 연구원. 1993년 Xerox Palto Altodusrnth 자문. 1997년 ~ 1999년 문체부 국어심의회 정보분과 위원. 관심분야는 한국어정보처리, HCI, 정보검색