

어휘적 중의성 제거 규칙과 부분 문장 분석을 이용한 한국어 문법 검사기 성능 향상

소 길자 부산대학교 전자계산학과 부산시 금정구 장전동 산30 kjs@solge.cs.pusan.ac.kr	남 현숙 부산대학교 전자계산학과 부산시 금정구 장전동 산30 hsnam@solge.cs.pusan.ac.kr	김 수남 부산대학교 전자계산학과 부산시 금정구 장전동 산30 snkim@solge.cs.pusan.ac.kr
원 상연 부산대학교 전자계산학과 부산시 금정구 장전동 산30 jupiter@solge.cs.pusan.ac.kr	권 혁철 부산대학교 전자계산학과 부산시 금정구 장전동 산30 hckwon@hyowon.cc.pusan.ac.kr	박 동인 한국전자통신연구원 대전 유성구 가정동 161 dipark@etri.re.kr

Improvement of Korean Grammar Checker Using Partial Parsing based on Dependency Grammar and Disambiguation Rules

KilJa,So Department of Computer Science Pusan National University SangYeon,Won Department of Computer Science Pusan National University	HyeonSook,Nam Department of Computer Science Pusan National University HyuckChul, Kwon Department of Computer Science Pusan National University	SuNam,Kim Department of Computer Science Pusan National University DongIn, Park Electronics and Telecommunications Research Institute
--	--	---

요약

한국어 문서에는 여러 어절을 검증해야만 처리할 수 있는 의미 오류와 문체 오류가 있다. 다수 어절 사이에 있는 오류는 부분 문장 분석을 한다. 논문에서는 의미 오류와 문체 오류를 처리할 때 어휘적 중의성 때문에 생기는 문제점을 제시하고 해결방법을 제안한다. 어휘적 중의성이란 한 단어가 두 가지 이상의 형태소 정보를 가짐을 뜻한다. 철자검사기와는 달리 문법 검사기에서는 어휘적 중의성을 제거하지 않으면 여러 가지 검사 오류가 발생한다. 이 논문에서는 의미, 문체 시스템에서 어휘적 중의성 때문에 검사 오류가 발생할 수 있는 과정을 크게 세 단계로 분류하였다. 연어 오류가 발생할 수 있는 검사단어가 어휘적 중의성을 가지면 표제어가 다른 규칙이 여러 개 존재한다. 이 때 규칙 선택 문제가 생긴다. 중의성 문제는 부분 문장분석 과정에서 지배소와 의존소 사이의 의존관계를 정확하게 설정하기 어렵게 한다. 본

논문에서는 각 단계에서 발생한 문제를 최소화하여 문법 검사기의 성능을 향상시킨다.

1 서론

한국어 문서에는 철자법 오류와 입력 오류 외에도 여러 어절을 검증해야만 처리할 수 있는 의미 오류와 문체 오류가 있다. 다수 어절에 걸친 검증이 필요한 오류들을 정확히 교정하려면 구문분석을 행해야 한다[9]. 그러나 구문 분석을 하면 구문 오류를 정확히 찾을 수는 있으나, 속도가 느리고 많은 기억 공간이 필요하므로 효율적이지 못하다. 그러므로 한국어 문서에서 자주 나타나는 오류 유형을 분석하여 규칙화하고 유형별로 분류하였다[6],[4]. 그리고 구문 분석이 필요한 규칙은 의존 문법을 이용한 부분 문장 분석으로 연어 오류 단어를 찾았다[4].

기존 시스템에서 사용한 부분 문장 분석 알고리즘은 검사단어를 지배소로 왼쪽 어절이나 오른쪽 어절을 읽어서 어절 분석을 한 후 어절간의 의존 관계를 보았다. 현재 의존소가 지배소나 검사단어와 의존 관계가 성립하지 않으면 비문으로 간주해서 문장 분석을 끝내고 의존 관계가 성립하면 현재 단어를 지배소로하고 다음 단어를 의존소로 해서 연어 오류 단어를 찾을 때까지 문장 분석을 한다[4]. 이 때 규칙베이스에서 적용할 규칙은 하나 이상일 수 있다.

따라서 각 규칙을 적용할 때마다 부분 문장 분석을 수행한다. 그러나 본 논문에서는 문장 분석한 결과를 저장한 후 다음 규칙을 적용할 때는 이미 분석된 결과를 이용하도록 한다.

한국어가 가지고 있는 중의성은 의미 오류와 문체 오류를 찾는 과정에서 여러 가지 오류를 일으킨다. 중의성에는 구조적 중의성과 어휘적 중의성이 있다. 문법 규칙에 따라 문장의 구조를 해석할 때 두 가지 이상의 구조로 해석되는 문장은 구조적 중의성이 있는 문장이다. 어휘적 중의성은 한 단어가 두 가지 이상의 형태소 열로 분석되는 현상을 뜻한다. 본 논문에서는 문법 검사기에서 어휘적 중의성이 발생시키는 문제점과 해결방법을 제안한다.

본 논문은 다음과 같은 단계를 거쳐 전개된다.

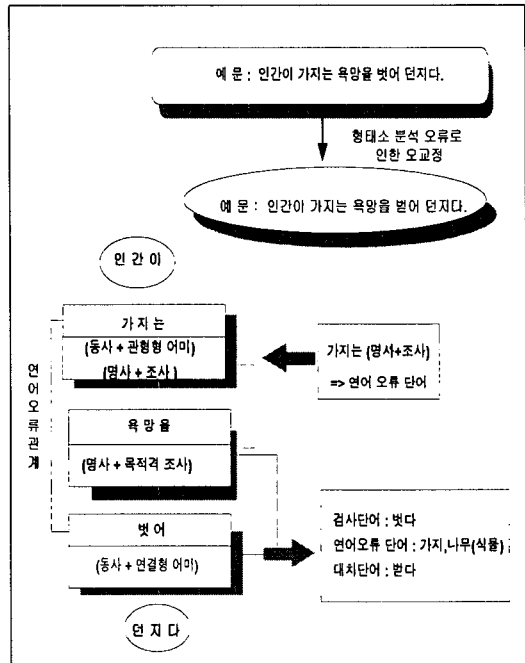
2장에서는 문법 검사기에서 어휘적 중의성 때문에 발생하는 문제점들과 중의성 해결이 요구되는 처리 단계를 설명하고, 3장, 4장, 5장에서 해결방법을 제시한다. 6장에서 결론을 맺고 향후 연구과제를 제시한다.

2. 문법 검사기에서 중의성에 의해 발생하는 문제점.

철자 검사기는 현재 어절의 오류 여부를 판단하는 시스템이기 때문에 어휘적 중의성을 고려하지 않아도 된다. 그러나 문법 검사기에서 어휘적 중의성을 제거 하지 않고는 정확한 검사가 불가능하다. 한국어에서 태깅은 형태소 분석결과와

중의성을 제거하는 방법으로 확률을 이용하는 방법, 퍼지망을 이용하는 방법, 신경망을 이용하는 방법 등 다양하다. 여러 개의 품사를 가진 단어에 대해서 하나의 품사를 결정하기 위해서 여러 분야에서 품사 태깅 시스템이 사용되고 있다. 그러나 현재 철자 검사기에는 품사 태깅 시스템이 부착되어 있지 않다. 본 연구에서는 품사 태깅 시스템 없이 의미 오류나 문체 오류를 검사하는 과정에서 휴리스틱 규칙을 이용하여 중의성을 제거하고자 한다.

다음은 문법검사가 의미 오류를 처리하는 과정에서 어휘적 중의성 때문에 잘못된 대치어를 생성하는 예다.



[그림 1] 중의성에 의해 생긴 교정 오류 예

'가진다'는 '명사+주격조사'와 '용언+관형형어미'로 해석할 수 있다. 의미, 문체 검사기가 '가진다'를 '명사+주격조사'로 해석하면 '가지'와 '벗다'는 연어 오류 관계기 때문에 '벗어'를 '받어'로 교정한다. 이 예문에서 '가진다'는 '명사+조사'보다는 '동사+관형형어미'로 보는 것이 옳다. 그러나 단어

단위 형태소 분석 정보만 이용할 때는 '가지는'의 올바른 형태소 분석 정보를 알 수 없다. 문법 검사기에서 이와 같은 어휘적 중의성 때문에 오류가 발생할 수 있는 과정은 크게 세 부분으로 나눌 수 있다. 철자 검사기에서 의미 문체 시스템을 호출할 때 넘겨주는 검사 단어의 어휘가 중의적일 때, 부분 문장 분석 할 때, 연어 오류 단어의 품사를 결정 할 때 어휘적 중의성이 제거되지 않으면 올바른 검사를 할 수 없다.

3. 검사단어의 어휘가 중의적일 때 RULE SELECTION 방법

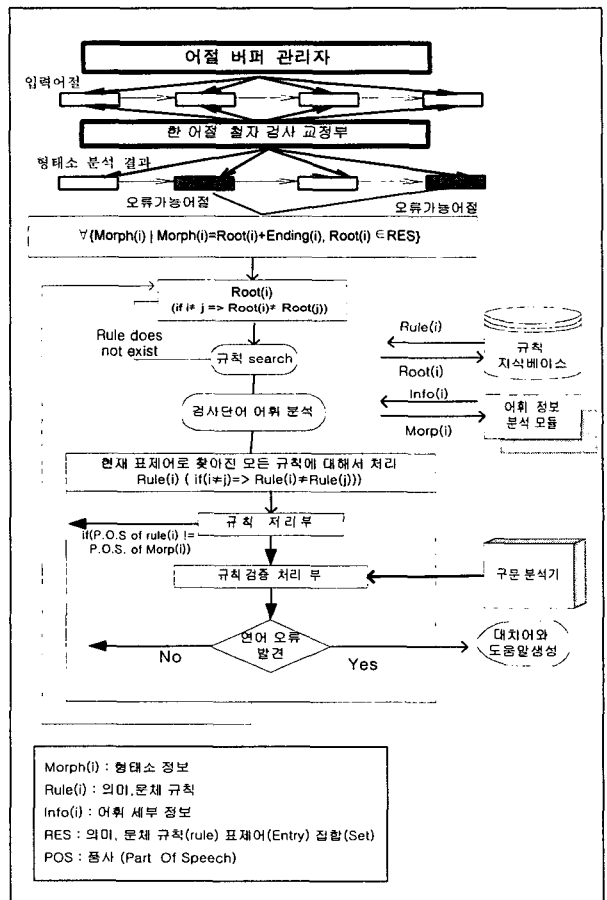
문법 검사기에서 의미 오류와 문체 오류를 검증하는 부분은 크게 규칙 검색부, 규칙 검증 처리부, 대치어-도움말 생성부로 구성되어 있다. 철자 검사기는 현재 어절을 형태소 오토마타를 따라 가면서 분석하다가 의미 오류나 문체 오류가 발생할 가능성이 있는 어절이 있으면 의미 문체 처리의 규칙 검색부를 호출한다. 규칙 검색부는 어절의 어간(Root)을 표제어로 규칙베이스에서 규칙을 찾는다. 이 때 철자 검사기는 현재 분석 중인 형태소 정보를 이용해 문법 검사기를 호출한다. 그러나 어휘적 중의성이 있는 어절은 어간의 형태가 다르게 분석될 수 있으므로 한 가지 형태소 정보만 이용하면 문맥상 적합한 규칙을 찾지 못할 수 있다. 본 논문에서는 이 문제를 해결하려고 검사단어의 형태소 정보 중 규칙의 표제어를 어간으로 하는 모든 형태소 분석 정보를 이용한다.

예를 들면 예제1에서 어절 버퍼 관리자가 철자 검사기에 '보고'를 넘겨주면 철자 검사기는 '보고'를 명사로 검사한다. 이 때 '보고(명사)'가 의미 오류를 발생할 표제어 SET에 있으므로 문법 검사기를 호출한다. 그러나 문법 검사기에 예제1에 있는 오류를 검사하려면 동사 '보다'의 어간인 '보-'를 표제어로 규칙을 검사해야 한다. 또 예제2에 있는 오류를 검사하려면 '보고(명사)'를 표제어로 규칙을 찾아야 한다

예제1 : 그를 올려다 보고 이야기를 했다 (x)
 (언어오류단어) (검사단어)
 => 그를 올려다보고 이야기를 했다 (0)

예제2 : 상황을 그에게 보고 했다 (x)
 (검사단어) (언어오류단어)
 => 상황을 그에게 보고했다 (o)

[그림 2]는 어휘적 중의성을 가진 검사단어에 의해 여러 개의 규칙이 존재할 때 의미, 문체 시스템에서 문맥상 적합한 의미, 오류 규칙을 선택하는 과정을 보여준다.



[그림 2] 문법 검사기에서 규칙 Selection 과정

- 1) 문법 검사기는 오류 가능 어절의 여러 가지 형태소 정보 중 어간이 규칙의 표제어 SET(Set of Rules'title)에 있는 형태소 분석 정보를 뽑아낸다.
- 2) 가장 긴 표제어를 어간으로 하는 형태소 정보를 먼저 이용해서 규칙을 찾는다. 하나의 표제어에는 의미, 문체 오류가 발생할 가능성이 있는 문장 패턴이 여러 개 존재할 수 있으므로 규칙베이스에서 찾아 적용할 규칙이 하나 이상일 수 있다[6].
- 3) 규칙 전처리기에서는 규칙들을 버퍼에 담고 입력된 검사 단어의 품사와 규칙에 설정된 검사 단어의 품사가 같은지 비교해서 같을 때만 규칙을 적용한다.
- 4) 규칙 검증 처리부는 왼쪽, 오른쪽 한 어절만 보거나 부분 문장 분석을 해서 규칙에 설정된 언어 오류 단어를 찾는다. 언어 오류 단어가 발견되면 대치어와 도움말을 제공한다.

[그림2]에 설명된 방법으로 했을 때 예제1은 '보'를 표제어로 찾은 규칙을 적용하여 언어 오류 단어를 찾는다. 예제2는 '보고'를 표제어로 찾은 규칙을 적용하여 언어 오류 단어를 찾는다. 문법 검사기는 예제1에서 '보다'와 언어 오류 단어인 '올려다'를 발견하고 '올려다보고'를 대치어로 생성한다. 예제2에서는 '보고'와 언어 오류 단어인 '하다'를 발견하고 '보고했다'를 대치어로 생성한다.

예제 3 : 이 사업은 이득이 매우 좋다. (x)
 (언어오류단어) (검사단어)
 => 이 사업은 이득이 매우 많다. (o)

예제3에서 검사단어 '좋다'는 '좋(형용사)+다'와 '좋다(감탄사)'로 분석된다. 예제3에서 사용된 '좋다'는 "이득'과 언어 오류 관계가 있는 형용사다. 예제3의 오류를 검사 및 교정하려면 '좋다'가 '좋+다(형용사+종결어미)'로 검사돼서 의미, 문체 오류 검사가 수행돼야 한다. 그러나 현재 철자검사는 '좋다'를 감탄사로 분석하고, 감탄사 '좋다'가 의미 오류 규칙베이스의 표제어SET에 없으므로 의미, 문체 오류 검사를 안 하고 검사를

마친다. 이 문제를 해결하려고 본 논문에서는 'Dummy Rule'을 사용한다.

'Dummy Rule'은 규칙이 저장되어 있지는 않지만 검사단어의 어간을 규칙 표제어 SET에 추가해서 철자 검사기가 문법 검사기를 호출할 수 있도록 하는 방법이다.

4. 의존 관계를 이용한 부분 문장 분석

[그림2]에서 규칙 검증 처리부는 의미/문체 오류를 검사하기 위해 검사 단어를 기준으로 왼쪽, 오른쪽에 언어 오류 관계가 있는 단어를 조사한다. 검사단어와 언어 오류 단어 사이에 여러 어절이 있을 때는 구문 분석기에서 의존 문법에 기반한 부분 구문 분석을 한다. 구문 분석기는 크게 어절 버퍼 관리자, 어휘 분석 모듈, 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. 중의성 제거 모듈은 중의성 제거 규칙을 이용해서 분석 중인 어절의 형태소적 중의성을 제거한다. 그리고 의존 관계 결정 모듈은 규칙에 의해 연속한 두 어절의 의존 관계를 조사한다. 의존 문법에서는 의존 가능한 두 가지의 어절 유형을 규정한 것이 문법규칙이 된다[2].

본 논문에서 의존 관계 규칙을 찾을 때 의존 관계가 성립하는 규칙보다는 의존 관계가 성립하지 않는 규칙을 이용해서 의존 관계를 결정한다. 즉 의존 관계 허용 범위를 넓혀서 언어 오류 단어를 찾을 수 있는 범위를 확대했다. 그리고 분석 범위가 너무 확대될 때 발생할 수 있는 속도 저하와 오류 발생을 막는 방법으로 문법 검사기가 검증할 수 있는 어절 수를 제한하거나 문장 분석을 종료하는 조건(Stop-Condition)을 둔다. 문장 분석 종료 조건은 용언 종결형이 나오거나, 필수 문장 요소인 목적어나 주어 다음에 '용언+연결어미'가 오거나 문장 종결 부호(!,?,...,)가 올 때다.

다음은 검사단어를 기준으로 왼쪽으로 구문 분석을 할 때 필요한 규칙이다.

지배소	의존소	예
용언	용언+관형형어미	착한 예쁘고
용언	용언+종결형어미	착하다 예쁜
용언	명사+소유격	나의 예쁜
용언	명사+접속격	학생과 예쁜
명사	명사+주격조사	피가 가슴에
명사	명사+목적격조사	사랑을 가슴에
명사	명사+부사격조사	행복으로 마음을
관형사	명사+조사	전문가가 여러
관형사	용언+용언연결형	아름다운 여러
부사	부사이외의 모든 품사	아름다운 매우
	...	

[표1] 왼쪽 방향 구문 분석에서 의존관계가 없는 어절관계

두 어절간의 의존 관계를 결정하려면 세부적인 형태소 정보가 필요하다. 본 논문은 구문 분석기에 입력된 어절을 다음과 같은 세부적인 정보를 이용하여 의존 관계를 결정한다.

품사	세부 형태소 정보	예
명사	명사 + 주격 조사	학생은, 학생도
	명사 + 목적격조사	학생은, 학생을
	명사 + '을/를'	학교를
	명사 + 소유격조사	학생의
	명사 + 접속격조사	학생과
	명사 + 부사격조사	학생에게
	명사 + '이다'축약 조사	공부다
용언	'하다' 결합 명사	공부하다
	'되다' 결합명사	처리되다
	'이다' 결합명사	공부이다
	동사 + 관형형어미	먹는
	동사 + 연결형어미	먹고
	동사 + 부사격전성어미	먹게
	동사 + 종결형어미	먹다

	형용사 + 관형형어미	예쁜
	형용사 + 연결형어미	예쁘고
	형용사+부사격전성어미	예쁘게
	형용사 + 종결형어미	예쁘다
술어 부사	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
기사	용언+ 어미	들려, 들리는데

[표 2] 의존관계 설정에 필요한 어휘 세부 정보

다음은 위에서 설명한 어휘 분석결과를 이용하여 의존 문법으로 검사단어의 왼쪽 또는 오른쪽 방향으로 구문 분석을 하는 알고리즘이다.

- 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-ALLI-NFO(WORD(I))
- 4) RMAMBIGUITY(MORP(I))
- 5) for (IndexOf Gov=I-1; IndexOfGov >=0;) {
- 6) If(MORP(IndexOfGov)와 MORP(I) 의존관계 성립){
- 7) GOV(I)= IndexOfGov
- 8) break;
- 9) }else
- 10) IndexOfGov = GOV(IndexOfGov)
- 11) IndexOfGov=NOT_EXIST
- 12) } // end for
- 12) STOREArray(WORD(I), MORP(I), Gov(I))
- 13) }else
- 14) IndexOfGov= RETRIVEArray(I)
- 15) if(IndexOfGov != NOT_EXIST) return TRUE
- 16) else return STOP_NOW

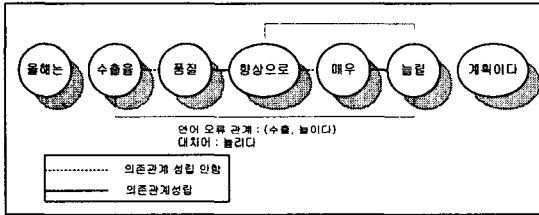
[그림 3] 구문 분석 알고리즘

(3)에서 어휘 분석을 하여 (4)에서는 형태소적인 중의성이 있으면 휴리스틱 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 문장 요소의 위치가 다르므로 구문 분석을 할 어절 수는 달라질 수 있다. 그러나 이미 의존 관계가 설정된 어절은 다시 분석하지 않는다.

다음은 위에서 설명한 의존 관계 규칙들과 알고리즘을 이용해서 부분 문장 분석을 하는 예를 보인다.

예제4: 올해는 수출을 품질 향상으로 매우 늘일 계획이다



[그림 4] 문장 분석 예

'늘이다'는 '더 길게 하다'는 뜻으로, '수출, 생산, 판매, 범죄, 수입, 의료비, 수량, 예산, 실적' 등과 연어 오류 관계가 있다. [그림 4]에서 검사 단어 '늘일'은 '주어, 목적어,' 용언 수식어' 등을 의존소로 가질 수 있다. 왼쪽 어절 '매우'는 부사로 용언을 수식할 수 있으므로 의존 관계가 성립한다. 그러나 '향상으로'와 '매우'는 의존 관계가 성립하지 않는다. 그러므로 '매우'의 지배소('늘일')와 의존관계를 살핀다. '향상으로'와 '늘일'은 용언 수식어와 용언의 관계가 있으므로 의존 관계가 성립하고 '향상으로'의 지배소가 '늘일'이 된다. '수출을'은 '품질'과

의존관계가 성립하지 않고, '향상으로'와도 의존 관계가 성립하지 않는다. '향상으로'의 지배소인 '늘일'과는 목적어와 서술어의 관계가 성립하므로 '수출을'의 지배소는 '늘일'이 된다. '늘일'과 연관 관계에 있는 문장 요소(목적어)가 발견되었으므로 부분 구문 분석을 끝내고 의미 오류 처리 규칙에 기반한 검사 교정을 수행한다.

5. 어휘적 중의성을 제거한 구문 분석기

문장 분석을 하다 보면 의존 관계를 결정할 어절의 어휘 정보가 중의적이어서 의존 여부를 결정하기 애매할 때가 있다. 본 논문에서는 두 가지 방법을 이용해 중의성을 제거 한다. 중의성 제거 규칙은 어휘중심 중의성 제거 규칙과 품사중심 중의성 제거 규칙으로 분류할 수 있다. 어휘중심 중의성 제거 규칙은 입력 문장의 각각의 어휘에 대해 규칙이 존재한다. 예를 들면 앞 어절의 어미가 '-어야', '-아야', '-해야'로 끝난 용언일 때 '하다'는 보조용언이다. 그리고 품사중심 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 구문 분석을 할 때 먼저 분석된 어절이 용언이고 왼쪽에 이웃한 어절이 '동사+관형형', '명사+목적어'로 중의적 어휘정보를 가질 때 '동사+관형형' 보다는 '명사+목적어'를 우선으로 선택한다.

5.1 어휘 문맥 정보를 이용한 중의성 제거 규칙

어휘 중심 중의성 제거 규칙은 다음과 같은 형태로 구성된다.

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

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+어미'와 '부사'로 해석될 때 중의성을 제거하는 규칙은 다음과 같다.

본용언 '하다'는 '정책'과 연어 오류 관계다.

RULE(1)=
 (하다, LEFT, e 어야 | e 아야 | e 해야, 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)

'정책을 하다'보다는 '정책을 퍼다'로 쓰는 것이 바람직하다. 그러나 예문5에서 사용된 '하다'는 왼쪽 어절의 어미가 '-아야'가 나왔기 때문에 보조 용언으로 해석되어 '정책'과 연어 오류 단어가 될 수 없다. 그러나 예제6에서는 '하다'가 본용언이고 연어오류단어 '정책을'이 발견되었으므로 '하는'을 '퍼는'으로 교정한다.

예제5 : 새로운 정책을 과거청산 방법으로 보아야 한다.

예제6 : 인간중심 정책을 하는 정치인 (x)
 (연어오류단어) (검사단어)
 => 인간중심 정책을 퍼는 정치인 (o)

5.2 품사 정보를 이용한 중의성 제거 규칙

이 규칙은 구문 분석을 할 때 이미 분석된 어절의 품사를 이용해서 그 다음 입력된 어절의 품사를 결정하는 방법이다. 즉 검사 단어를 시작으로 구문 분석 방향에 따라 입력된 어절의 어휘가 중의적일 때 앞에 분석된 어절의 품사를 이용해서 중의성을 제거하려는 방법이다. 이 방법은 이미 앞 단계에서 분석된 정보를 이용해 현재 단어의 중의성을 제거하기 때문에 구문 분석 과정에서 중의성 제거에 필요한 시간을 줄일 수 있다.

품사 중심 중의성 제거 규칙은 다음과 같은 형태로 구성된다.

RULE(#)=((Left,Condition.Col/Anti),
 (Right, Condition,Col/Anti), Selection,)

중의성 유형은 중의적으로 분석된 형태소 정보를 나타내고 condition은 중의성을 제거하는데 필요한 왼쪽 또는 오른쪽 어절의 품사 정보를 나타낸다. Anti_collocation은 왼쪽이나 오른쪽 어절이 Condition에 명시된 품사가 아닐 때 조건을 만족하고, Collocation은 왼쪽이나 오른쪽 어절이 Condition에 명시된 품사로 분석될 때 조건을 만족한다, 조건이 만족되면 규칙에 의해 Selection에 명시된 품사를 선택한다.

예로 '명사+ 조사'와 '용언+관형형어미'로 해석되는 어절에서 중의성을 제거하는 규칙은 다음과 같다.

RULE(1) ((Left,0,0),(Right, N, Anti), (N+J)|N);
 RULE(2)
 ((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+조사'나 '조사가 생략된 명사'로 본다. 오른쪽 어절이 명사고, 왼쪽 어절이 '명사+소유격조사'나 '명사' 또는 '용언+관형형어미'가 아닐 때 현재 어절을 용언+관형형어미로 본다..

예제7에서 '불을'은 '명사+목적격조사'와

검사단어 : 부치다
 연어 오류 단어 : 불, 감시원,
 꼬리표, 마음, 말 등
 대치단어 : 불이다

예제7: 어머니가 화로에 불을 짚이고 켜셨다.
 (N+jco); (P+etm)
 => 어머니가 화로에 불을 붙이고 켜셨다

'동사+관형형어미'로 해석된다. 그러나 '짚이다'가 용언이기 때문에 규칙 RULE(1)에 의해서

'명사+목적격조사'로 해석되고 '깊이다'와 언어 오류 단어가 된다. 문법 검사기는 언어 오류 관계 단어가 발견되었기 때문에 '깊이다'를 '지피다'로 교정하고 도움말을 제시한다.

[그림1]에서는 '가지는'이 '동사+관형형어미'와 '명사+주격조사'로 분석되어 검사 및 교정 오류가 발생했다. '가지는'의 뒤는 '명사+조사(욕망어미)'이고, 앞 어절이 '명사+주격조사'로 분석되었으므로 RULE(2)에 의해서 '가지는'은 '동사+관형형어미'로 해석하여 중의성을 제거한다. '가지는'이 '용언+관형형어미'로 해석되면 구문 분석 과정이 종료되므로 검사 오류가 발생하지 않는다.

6. 결론 및 향후 연구 과제

본 논문에서는 하나의 검사단어에서 표제어가 다른 여러 개의 규칙이 존재할 때 문맥상 적용해야 할 적합한 규칙을 찾는 방법을 제안했다. 그리고 중의성 제거 규칙을 이용하여 문법 검사기 수행과정에서 발생할 수 있는 오류를 제거하였다.

향후과제로는 한 검사 단어의 두 가지 규칙에서 동시에 언어 오류 단어를 찾았을 때 발생하는 Rule Conflict 문제를 해결해야 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.

참고문헌

- [1] 박용욱, 조혁규, 권혁철, '의존 문법을 이용한 한국어 분석기의 구현', '90정보과학회 봄 학술발표논문집, pp191-194, 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, 박용욱, 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 학술발표논문집, 1991

[8] 이하규, '어말-어두 공기 정보를 이용한 한국어 어휘 중의성 해소', 한국정보과학회논문지(B) 제24권 제1호, pp. 82-89

[9] 채영숙, '언어 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