

한국어 문법검사기에서 의미정보를 이용한 복합명사의 분석제약

*원상연, *김수남, *김광영 *남현숙, **권혁철
*부산대학교 전자계산학과, **부산대학교 정보·컴퓨터공학부
{jupiter, snkim, kykim, hsnam}@solge.cs.pusan.ac.kr
hckwon@hyowon.pusan.ac.kr

To Constrain Korean Compound Nouns using Semantic Information for Korean Grammar Checker

Sang-Yun Won, SuNam Kim, Kwang-young Kim, Hyunsuk Nam, Hyuk-Chul Kwon
Department of Computer Science, Pusan National University
Division of Information and Computer Engineering, Pusan National University

요 약

일반적으로 두 개의 명사가 결합하여 하나의 명사 기능을 하는 어구를 복합명사라고 한다. 한국어는 복합명사 내의 명사를 붙여 쓸 수도 있고 띄어 쓸 수도 있으므로 복합명사의 형태적 분석에 많은 어려움이 있다.

이 연구에서는 각 명사의 복합명사 결합을 최대한 제약하여 문법검사기에서 복합명사와 관련된 오류의 발생을 최소화할 수 있는 방법을 개발했다. 이 논문에서 복합명사 분석 기능을 제약하는 방법으로 형태적 제약 방법과 의미정보에 따라 복합명사의 결합관계를 제약하는 방법을 이용했다. 어휘 정보만으로 복합명사를 분석하면 의미관계에 의한 오류는 찾기 어려우므로 복합명사의 구조적 결합관계와 의미 결합관계를 밝혀 복합명사를 잘못 분석하는 문제점을 극복한다. 복합명사의 결합제약은 명사의 왼쪽과 오른쪽에 올 수 있거나 올 수 없는 명사들의 의미, 형태적 특성과 명사가 나타나는 분포(distribution)에 따라 분류하여 규칙베이스화하였다. 의미정보를 이용한 복합명사 결합제약 알고리즘도 구현하였다.

1. 개 요

복합명사의 정의는 단순히 두 명사가 이웃할 때 복합명사로 보는 견해부터 복합명사로 결합하여 뜻이 바뀌었을 때까지 다양하다. 일반적으로 두 개 이상의 명사가 결합하여 하나의 명사 기능을 하는 어구를 복합명사라고 한다[1]. 본 연구에서는 한 어절 내에 다수의 명사를 이웃하여 쓰는 어절을 복합명사로 보았다. 문법검사기, 기계번역, 정보검색 등 자연언어처리 응용시스템의 성능향상을 위한 연구에서 복합명사의 정확한 분석은 매우 중요하다. 한국어는 복합명사를 이루는 명사들을 붙여 쓸 수도 있고 띄어 쓸 수도 있으므로 복합명사의 형태적 분석은 복합명사가 많이 사용되는 문서(신문이나 잡지 등)에서 자연언어처리 응용시스템의 성능을 향상하기 위해서는 복합명사의 분석이 중요하다.

그러나 복합명사는 조합이 자유롭고 길이에 제한이 없어 분석에 어려움이 따른다. 특히 접사나 조사가 복합명사의 구성요소로 해석되는 문제는 복합명사 분석을 더욱 어렵게 한다. 이러한 복합명사 분석의 모호성을 제거하는 방법으로 사

전을 이용한 방법과 통계적 방법을 들 수 있다[2, 3, 4]. 통계정보를 이용한 방법은 말뭉치에 있는 단어의 빈도를 사람이 수동으로 분석해야 하므로 시간적·경제적 손실이 크다[5]. 따라서 본 논문에서는 하위 범주화된 어휘정보를 가진 전자사전을 이용하여 복합명사를 분석하였다. 그러나 어휘 정보만으로 분석했을 때 형태적으로 문제는 없지만 의미상으로 잘못된 분석이 발생할 수 있다. 본 논문에서는 의미 제약을 통해 틀린 복합명사를 맞게 분석하는 문제점을 극복한다. 복합명사의 결합제약은 명사의 왼쪽과 오른쪽에 올 수 있거나 올 수 없는 명사들의 의미·형태적 특성과 명사가 나타나는 분포(distribution)에 따라 분류하여 규칙베이스화하였다. 그리고 분석된 결과 구성요소끼리 복합명사를 이루지 못할 때 오류의 유형을 분석하고 유형에 따라 대치어를 생성한다.

본 논문에서 명사의 복합명사 결합기능을 제약하는 방법으로 형태적 제약방법과 의미적 제약방법을 이용했다. 각 명사의 복합명사 생성을 최대한 제약하여 문법검사기에서 복합명사와 관련한 오류의 발생을 최소화할 수 있는 방법을 개발했다.

2. 복합명사 결합제한

복합명사는 관점에 따라 다양하게 정의된다. 본 논문에서는 다수의 명사를 이웃하여 쓰면 이를 복합명사로 보았다. 이러한 복합명사는 조합이 자유롭고 길이에 제한이 없어 문법검사기의 한계로 작용한다.

복합명사 처리의 한계점
철자가 틀린 단어가 단순히 형태적 조합에 의해 복합명사로 분석됨으로써 문법검사기의 오류 검증 정확도를 떨어뜨린다. (예) 자연과학(×) → 자연과학(○) 고막요리(×) → 꼬막요리(○) 오존농도(×) → 오존농도(○)
오류단어에 대치어를 생성할 때 의미상 허용할 수 없는 복합명사를 생성하여 제시함으로써 시스템에 대한 신뢰도를 떨어뜨린다. (예) 프레온가스나(○) → 프레온계집아이(×)
띄어쓰야 의미를 쉽게 파악할 수 있는 명사를 붙여 씌으로써 글을 읽기 어렵게 한다. (예) 흡연여성야채기피(×) → 흡연여성 야채기피(○)

복합명사 분석오류의 종류는 첫째, 옳은 어절에 대해 분석이 실패한 예(1종 오류)와 둘째, 틀린 어절을 옳다고 분석하는 예(2종 오류)가 있다. 1종 오류는 옳은 어절을 틀린 어절로 분석하여 틀린 대치어를 제시하여 문법검사기의 신뢰도를 떨어뜨린다[표 1].

옳은 복합명사 어절	잘못된 대치어 생성
프레온가스나	프레온계집아이
배선공사	배선 공사
미라주전투기가	미라 주전투기
우전제비오름에	우전제비울음에
추가피해사실이	추가피하사실이
대통령직인수위	대통령직인 수위

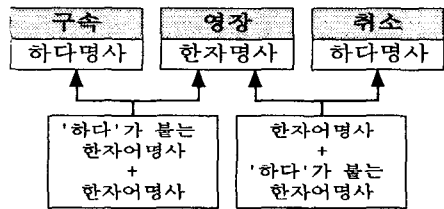
[표 1] 1종 오류의 예

2종 오류는 틀린 어절을 맞는 어절로 분석하여 문법검사기가 정확한 대치어를 제시하지 못해 오류검증의 정확도를 떨어뜨린다. 그리고 띄어쓰야 의미를 파악하기 쉬운 복합명사는 띄어쓰야 한다[표 2].

틀린 복합명사 분석	올바른 대치어 생성
상당+수조+직원	상당수 조직원
국가+도발+전하+고	국가도 발전하고
호주+산소+고기	호주산 소고기
발전+소경+비소	발전소 경비소
무선+호출+기용+반도체	무선호출기용 반도체
해수욕+장내	해수욕장 내
등기+부상+소유자	등기부상 소유자
전나무+의향	전나무의 향
대행업+체중	대행업체 중
자전+거위	자전거 위
돼지시기	돼지출하시기
생산돼지	생산한 돼지, 생산되지
어머니기지	어머니 기지
흡연여성야채기피	흡연여성 야채기피
상습성폭행	상습성 폭행 상습 성폭행
시인이자정치가	시인이자 정치가

[표 2] 2종 오류의 예

일반적으로 복합명사는 하나의 문장처럼 구성요소 사이에 어휘·의미적 관계가 존재한다. 우리말 복합명사는 병렬구조를 이루며 첫 번째 구성요소가 뒤에 오는 단어를 한정하는 기능을 한다. 이 구성요소 사이에는 어휘·의미적 언어관계가 성립한다.



[그림 1] 어휘 정보를 이용한 복합명사 분석

[그림 1] '구속영장취소'에서는 '하다'가 붙는 한자어명사와 한자어명사가 형태적으로 언어관계가 성립하므로 복합명사로 분석할 수 있다. 그러나 형태적 관계만 고려할 때 언어관계가 성립하나 의미상으로 맞지 않는 분석이다.

아래의 [표 3]은 어휘 정보만으로 분석했을 때 발생하는 잘못된 분석의 예이다.

복합명사	분석
국가도발전하고	국가 + 도발 + 전하 + 고
대형업체증	대형업 + 체증
애국심과자부심	애국심 + 과자 + 부심

[표 3] 잘못 분석한 복합명사의 예

이 논문에서는 복합명사 결합을 제약하는 방법으로 형태적 제약방법과 의미 정보를 이용한 제약방법을 쓰고 있다. 형태적 제약 방법은 사전의 어휘정보와 규칙을 기반으로 특정명사가 복합명사를 못 만들게 하거나 다른 명사 앞에서만 복합명사를 만들게 하는 방법이다. '지명'이나 '인명'은 원칙적으로 앞에 명사와 복합명사를 만들지 못하게 한다(예 : 연방하원의원김창준(×) → 연방하원의원 김창준(○)). 이 규칙은 기본적인 형태적 제약규칙의 예이다. 또한 '말기', '여유', '사실' 등의 명사는 어두에 올 때 다른 명사의 앞에 붙여 쓸 수 있지만 다른 단어의 뒤에 올 때 반드시 띄어 써야 한다[표 4].

해당 단어	사용 예	
말기	말기중세 말기환자	조선시대 말기 세포분열 말기
여유	여유공간 여유생활	생활 여유공간
사실	사실여부 사실판정	피해 사실

[표 4] 형태적 제약규칙의 예

'말기'와 유사한 의미로 앞 명사와 띄어써야 하는 명사의 예는 아래와 같다.

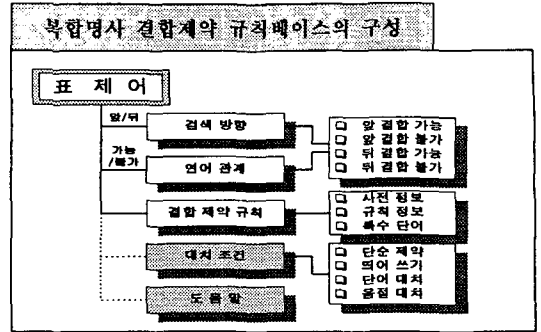
(예) 거대, 국내, 극렬, 농공, 농수산, 단식, 복식, 대내외, 대단위, 범민주, 비동맹, 비핵, 적극, 적당, 전국, 정초, 중기, 최고, 최근, 최대, 최상, 최신

의미 정보를 이용한 제약은 명사 간에 발생할 수 있는 의미적 결합정보를 체계화하여 규칙베이스를 구축하였다. 이 규칙베이스를 기반으로 하여 복합명사를 검사/교정할 때에 오류가 발생할 수 있는 복합명사의 결합을 제약한다('이전'은 왼쪽으로 "하다"와 결합 가능한 한자어명사"와 결합할 수 없다. 실시이전에(×) → 실시 이전에(○)).

3. 의미 정보에 따른 결합제약 규칙베이스 구축

결합제약은 중성이 없는 음절로 구성된 명사나 잘못 쓸 가능성이 큰 명사에 대해서 적용한다. 이웃하는 명사 사이의 관

계를 분석해서 세 개 이상의 명사가 결합된 예로 확장하여 분석할 수 있다.



3.1 의미 결합정보의 표제어 선택

표제어는 복합명사 결합오류를 발생시킬 수 있는 단어를 말한다. 표제어의 선택은 아래의 기준에 따라 선택했다.

- 표제어의 선택**
1. 많이 사용되는 단어
 2. 중성이 없는 음절로 구성된 명사나 잘못 쓸 가능성이 큰 단어
 3. 위치적 제약을 가지는 단어

모든 명사에 대해 복합명사 의미적 결합제약을 하는 것은 어렵다. 많이 사용되는 명사 중 중성이 없는 음절로 구성된 명사나 잘못 쓸 가능성이 큰 명사에 대해서 좌·우 연관 선호도 (Associative Preference)를 가지고 결합방향에 대한 제약을 줄 수 있는 단어를 선택함으로써 효과를 얻을 수 있었다.

3.2 검증방향 (결합제약 검증방향)

의미 검사 후보 명사를 중심으로 왼쪽에 오는 명사에 제약이 적용되는지, 오른쪽에 오는 구성요소에 적용되는지를 표시한다. 왼쪽 우선 연관(Left Association)의 비율이 70%정도를 차지한다[10]. 또한 명사가 앞에 올 때와 뒤에 올 때 의미가 서로 다른 예가 발생하므로 순서(결합방향)를 고려하여야 한다. 예를 들어 '신문판매'는 복합명사로 사용되지만 '판매신문'이라는 복합명사는 사용되지 않는다.

3.3 언어관계 (결합 관계)

표제어와 결합된 명사가 특정 단어와는 결합 가능한 언어관계 (Collocation relation)인지, 특정 단어와 결합 불가능한 언어 오류 관계(Anti-Collocation relation)인지의 여부로 단어 간에 결

합 가능 규칙인지 아니면 결합 불가능 규칙인지를 결정한다.

3.4 결합 제약 규칙

연어관계 정보는 어휘정보나 범주화된 의미정보(예: 사람, 동물, 식물, '하다'와 결합할 수 있는 명사 등), 개별단어를 이용하여 복합명사의 결합을 제약하는 규칙을 나타낸다. 규칙베이스에는 4,316개의 표제어와 114개의 범주정보를 가지고 명사 간의 의미적 결합을 제약한다. 범주정보는 명사의 형태적 특성과 의미에 따라 복합명사의 결합을 제약하려고 분류했다. [표 5]는 범주화된 의미정보의 일부이다.

범주정보	설명	예
사람명사	사람을 나타내는 한자어, 순수명사 등	어머니, 아버지, 경찰관
동물명사	포유류, 조류, 파충류, 균류, 곤충	호랑이, 사슴, 돼지, 말
식물명사	식물 명사	장미, 소나무
'하다'명사	'하다' 결합이 가능한 명사로 동작성 / 상태성을 표시	청구, 판매, 할인, 취소, 연기
지역명사	국가명, 도시명 등	한국, 서울, 부산
외래어명사	외래어	프로, 아노미

[표 5] 규칙베이스의 분류정보 예

3.5 대치어 생성규칙

대치조건은 연어관계가 성립하지 않을 때 대치어 생성방법을 설명한다. 검사모듈이 의미적 연어관계를 검사한 후 연어관계가 성립하지 않으면 규칙에 정의된 대치 방법으로 대치어를 생성한다[표 6].

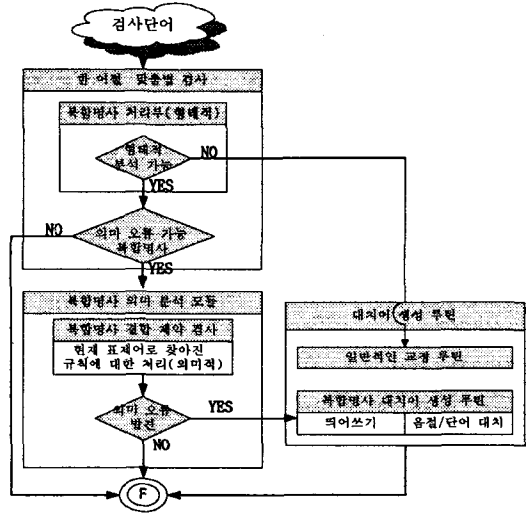
대치어 생성 규칙	설명
단순 제약 (형태적)	형태적으로 복합명사가 될 수 없는 단순한 명사 간의 결합이다. 맞춤법 검사/교정기의 교정 모듈을 이용하여 교정한다. (예) 기계화영농(×) → 기계화 영농(○)
단어 간 띄어쓰기 (의미적)	현재 오류 가능 명사가 포함된 복합 명사는 이 오류 가능 명사를 중심으로 띄어쓰기 교정을 한다. (예) 대통령취임이전(×) → 대통령취임 이전(○)
음절 / 단어 대치 (의미적)	현재 오류 가능 명사는 복합 명사를 이룰 때 특정 음절/단어로 대치하여 의미를 올바르게 한다. (예) 공부하녀(×) → 공부하여(○)

[표 6] 대치어 생성규칙

4. 복합명사 의미 분석 시스템

4.1 시스템 구성

앞 절에서 언급한 규칙베이스를 적용한 형태소 분석기의 복합명사 분석모듈은 [그림 2]와 같다.



[그림 2] 복합명사 의미 분석 시스템

어떤 어절이 복합명사로 분석되었고, 규칙화된 명사가 복합명사 생성에 참여했으면 '복합명사 의미 분석 모듈'이 호출된다. 호출된 분석 모듈은 현재 어절의 구성요소에 특정단어의 의미상 연어관계를 검사할 필요가 있으면 복합명사 결합제약 검사모듈을 호출한다. '복합명사 의미 분석 모듈'은 규칙베이스에서 의미 검사 후보 명사의 결합 제약규칙을 찾아 이를 바탕으로 좌·우 명사와 연어 관계를 검사한다. 규칙베이스에는 각 명사에 대한 연어관계 정보, 검증방향, 결합 제약 규칙과 대치어 생성 규칙이 있다. 이때 연어 관계가 성립하면 복합명사가 가능함을 알리고 성립하지 않으면 규칙에 정의되어 있는 대치방법을 적용시켜 대치어를 생성한다. 복합명사의 결합제약 검사단계는 아래와 같다.

복합명사 결합제약 검사단계

[단계 1] 형태소 분석(형태적 제약방법)
어휘정보와 규칙을 통한 형태적 분석

[단계 2] 의미적 결합제약 검사
[단계 1]에서 형태적으로 올바른 어절로 분석된 복합명사 중 오류의 가능성이 있는 명사 간의 의미정보를 이용한 복합명사 결합 제약검사를 수행한다.

[단계 3] 대치어의 생성
대치어가 생성은 일반교정모듈과 복합명사 결합 제약 모듈에서 이루어진다. 대치의 방법에는 띄어쓰기, 단어대치 등이 있다.

[단계 1]에서 복합명사에 대한 맞춤법 검사 과정에서 복합명사를 구성하는 각 명사의 어휘정보를 이용하여 오류 여부를 분석한다. 8음절로 구성된 '흡연여성야채기피'는 '흡연'+ '여성'+ '야채'+ '기피'의 결합형태이다. 먼저 사전을 최장일치법으로 탐색한다. '흡연'은 어휘정보로 "하다"와 결합 가능한 한자어명사"정보를 가진다. 어절에서 분석되지 않은 어절 '여성야채기피'는 사전 탐색을 통해 '여성'은 '한자어로 된 사람명사' 정보를 얻는다. '하다명사'와 '사람명사' 간에는 형태적 제약조건이 없으므로 형태적 검사 방법으로는 오류가 없다. 이렇게 형태소 분석만으로는 "흡연여성야채기피"는 올바른 복합명사로 분석된다.

[단계 2] 형태소 분석으로 오류는 없으나 복합명사 결합 오류를 발생시킬 수 있는 특정단어의 의미적 제약조건을 이용하여 추가의 검사 및 교정이 필요하다. 위의 예에서 '여성'과 '야채'는 형태소 분석에 따르면 '사람명사'와 '한자명사'의 형태적 결합으로 오류는 없으나, '야채'가 의미분석을 요구하는 명사이므로 의미관계 분석을 해야 한다. '야채'의 의미 규칙은 왼쪽 방향으로 '사람명사', '동물명사', '식물명사'는 올 수 없다. 그래서 '여성'과 '야채'는 의미적으로 복합명사로 분석될 수 없다.

[단계 3]에서 복합명사 대치어 생성 모듈을 통해 '흡연여성야채기피'는 '흡연여성 야채기피'로 의미를 쉽게 파악할 수 있게 띄어쓰기를 한다.

4.2 대치어 생성

대치조건은 언어관계가 성립하지 않을 때 대치어 생성방법을 설명한다. 맞춤법 검사모듈에서 의미상 언어관계를 검사한 후 언어관계가 성립하지 않으면 규칙에 정의된 대치 방법으로 대치어를 생성한다. 대치어를 생성하는데 그 방법은 형태적 분석과 의미적 분석과정으로 나누어 처리한다. 형태적 분석방법에서는 맞춤법 교정모듈을 통하여 대치어를 생성한다[표 7].

잘못 쓴 복합명사	대치어	형태적 제약
연방하원의원김창준	연방하원의원 김창준	인명, 지명
삼권분리, 권력분리	삼권분립, 권력분립	오용어
민주 주의	민주주의	절미사
국가상호, 학생상호	국가 상호, 학생 상호	다른 명사의 앞에만 쓰임
도시개발특별자금	도시개발 특별자금	
공장최신자동화	공장 최신자동화	
개발대상전체면적	개발대상 전체면적	

[표 7] 단순제약 대치의 예(형태적)

의미적 분석과정에서는 규칙에 제시된 대치조건에 따라 생성한다. 대치조건은 크게 2가지로 정의된다. 첫째, 단순히 구성요소를 띄어쓴다[표 8]. 둘째, 구성요소를 음절· 단어대치를 한다[표 9].

잘못 쓴 복합명사	대치어	결합 제약 규칙
일부당직자	일부 당직자	'일부'+사람명사(X)
대행업체중	대행업체 중	하다명사+'체중'(X)
출국자수 노동자수	출국자 수 노동자 수	하다명사+'자'가 사람명사 (예외) 동양자수
전문인수	전문인 수	하다명사+'인'이 사람명사
사업가수	사업가 수	하다명사+'가'가 사람명사 (예외) 노동가수
전나무의향	전나무의 향	식물명사+'의향'→'의 향'

[표 8] 띄어쓰기 대치의 예 (의미적)

잘못 쓴 복합명사	대치어	결합 제약 규칙
청구해고	청구하고	하다명사+'해고'→'하고' (예외)정리해고
청구하구 정리하구	청구하고 정리하고	하다명사+'하구'→'하고'
청구회수	청구횟수	하다명사+'회수'→'횟수' (예외) 분할회수, 투자회수
한강도화	한강도하	지역명사+'도화'→'도하'
인천연한	인천연안	지역명사+'연한'→'연안'
부산번지	부산주소	지역명사+'번지'→'주소'
독립하녀	독립하여	하다명사+'하녀'→'하여'
장미사육	장미재배	식물명사+'사육'→'재배'
사슴납용	사슴납획	동물명사+'납용'→'납획'
호랑이인구	호랑이숫자	동물명사+'인구'→'숫자'
어머니이식	어머니의식	사람명사+'이식'→'의식'
간호학작사 물리학작사	간호학박사 물리학박사	"학/론/법/주의/사"로 끝나는 한자명사+'작사' →'작곡'

[표 9] 단어·음절대치의 예(의미)

5 실험 및 평가

의미 정보를 이용한 복합명사 결합제약의 성능 평가를 위해 신문, 말뭉치 등의 각종 문서에서 무작위로 추출한 125,839개의 복합명사를 사용하였다.

5.1 "어휘정보" vs "어휘정보 + 의미정보"

본 논문에서 제시한 복합명사 결합제약 성능 평가를 위해 어휘 정보만으로 처리한 복합명사 분석과 의미 정보를 도입한 분석의 성능 비교실험을 하였다. [표 10]에서 어휘 정보만으로 복합명사를 분석했을 때 실험 어절 125,839개 중 6,845개의 복합명사로 잘못 인식하는 어절을 찾았고 반면 어휘 정보와 의미 정보를 동시에 이용했을 때 7,759개의 복합명사로 잘못 인식하는 어절을 찾았다.

	복합명사로 분석된 것	복합명사로 잘못 분석된 것
형태소	118,994	6,845
형태소+의미	118,080	7,759

[표 10] 실험 결과

어휘 정보만 가지고 분석했을 때 틀린 복합명사를 검사하는 비율이 낮은 이유는 의미상으로 분석이 불가능한 틀린 복합명사를 옳은 복합명사로 잘못 분석했기 때문이다. 틀린 어절을 맞는 복합명사로 분석한 오류(2중 오류)에 대한 검사율을 13.5% 향상했다.

5.2 대치어 생성 분석

어휘정보와 의미정보를 이용한 복합명사 분석에서 복합명사로 잘못 분석된 어절이 7,759개 발견되었다. 이 어절에 대해 8,025개의 대치어가 생성되었다. 띄어쓰기 74.3%, 음절·단어대치 13.7%, 복합명사 의미적 결합제약 9.1%로 대치어를 생성했다[표 11].

	형태적 제약방법 + 의미정보	
대치어 없음	223 (2.9 %)	
띄어쓰기	5,966 (74.3 %)	
단어·음절 대치	1,104 (13.7 %)	
복합명사 결합제약 에 의한 대치어 생성	전체	732 (9.1 %)
	띄어쓰기 대치	729
	단어·음절 대치	3
Total	8,025 (100 %)	

[표 11] 대치어 생성 분석

6. 결론

본 논문은 특정단어의 의미 정보를 이용하여 복합명사의 의미적 연어관계를 분석하였고, 틀린 복합명사를 맞게 분석하는 문제점을 극복했다. 또한 분석 실패 시 올바른 어절로 교정하는 방법을 제시하였다. 규칙 베이스를 기반으로 복합명사가 과도하게 생성되는 것을 억제하고 잘못된 복합명사를 교정하는 알고리즘도 구현하였다. 형태적 정보만을 가지고 복합명사를 검사하는 기존의 맞춤법 교정 시스템의 한계를 극복하였다. 제시된 방법의 장점은 다음과 같다.

첫째, 의미적 연어관계를 분석하여 복합명사 분석의 정확성을 높였다. 실험에서 복합명사를 이룰 수 없는 단어의 생성제약은 어휘 정보만 사용한 분석보다 13.8% 높였다.

둘째, 복합명사 분석 실패 시 올바른 어절로 교정하는 방법을 제시하였다. 분석에 실패한 복합명사 어절에 대해 올바른 대치어와 도움말을 제시하여 복합명사 사용에 도움을 주었다.

셋째, 의미 정보를 이용한 복합명사 결합제약을 통하여 복합명사로 결합했을 때 오류를 발생시킬 수 있는 명사에 대해서만 오류를 검사하여 시스템 전체 속도를 떨어뜨리지 않고 성능을

향상시켰다.

향후 연구 과제로는 특정단어의 의미 제약조건을 의미정보 별로 더욱 세분화하여 분류할 것이다. 규칙화한 규칙베이스를 자동으로 구축하는 방법과 미등록어나 한자가 합성된 복합명사의 처리 방법을 연구하고자 한다.

7. 참고문헌

- [1] 이기문 감수, “동아 새국어사전”, 동아출판
- [2] 김지영, 권혁철, “합성명사의 의미 관계 분석 시스템을 위한 지식베이스 구축 기법”, 한국정보과학회 가을 학술발표논문집, 1992, pp.985~968
- [3] Chae Young-Soog, Dong-In Park, Hyuk-Chul Kwon, “Analysis of Korean Compound Noun Based On Statistics Collected from Corpus”, The 17th International Conference on Computer Processing of Oriental Languages (ICCPOL97), Vol.11, 1997, pp.709~714
- [4] 윤보현, 임희석, 임해창, “통계정보를 이용한 한국어 복합명사의 분석방법”, 한국정보과학회 봄 학술대회논문집, 제22권, 1호, 1995, pp.925~928
- [5] 박혁로, 신중호, “비터비 학습 알고리즘을 이용한 한글 복합명사 분석”, 한국정보과학회 가을 학술대회논문집, 1997, pp.219~222
- [6] Makoto Nagao, “자연언어처리”, 홍릉과학출판사, 1998
- [7] 채영숙, 권혁철, “말뭉치로부터 추출된 통계 정보를 활용한 한국어 복합명사 분석”, 한국인지과학회 논문집, Vol.8, No. 2, 1997, pp.101~108
- [8] 권혁철, 채영숙, 김재원, 김민정, “한국어 철자 검색을 위한 형태소 분석 기법”, 우리말 정보화 큰잔치, 국어정보학회, 1991, pp.179~186
- [9] 최재혁(1996), “음절수에 따른 한국어 복합 명사 분리 방안”, 한글 및 한국어 정보처리, 1996, pp.262~267
- [10] 강승식, “한국어 형태소 분석을 위한 복합 명사의 인식방법”, 한글 및 한국어 정보처리 학술대회 논문집, 1996
- [11] 양성현, 송만식, “명사의 연어 정보와 서술성 명사의 공기공기정보를 활용한 복합명사 분석 및 자동색인”, 한글 및 한국어 정보처리 학술대회 논문집, 1996, pp.59-64
- [12] 김민정, 권혁철, “한국어 특성을 이용한 자동 색인 기법”, 한국정보과학회 가을 학술발표논문집 Vol.19, No. 2, 1992
- [13] 김동일, 이신원, 윤후병, 장세우, 정성중, “한국어 텍스트에서의 복합 명사구 출현 빈도에 관한 연구”, 한국정보과학회 봄 학술발표논문집 Vol.21, No. 1, 1994
- [14] 김지영, 권혁철, “지식에 기반한 한국어 복합명사의 의미 관계 분석”, 한국인지과학회 춘계 학술발표논문집, pp.113~129, 1992