

# 명사 어형의 생성과 그 조건

이기용(고려대학교 언어학과)\*

## Abstract

동사 어간과는 달리 명사 어간은 그 자체로도 어형(word form)이 될 수 있다. 그러나 일반적으로는 명사 어간에 조사가 하나 또는 여러 개가 결합되어 어형이 형성된다. 이 논문은 이러한 명사 어형을 효율적으로 생성할 수 있는 규칙기반의 어형 생성 시스템 KORMOR(KORean MORphological system)이 어떻게 운용되는가를 시연하는 것이 그 목적이다.

이 시스템은 변이형을 포함한 기본 조사 108개를 기초로 하여 3,000여개의 복합형 조사를 생성한다. 그러나 국어의 명사 어간에는 조사가 하나뿐 아니라 6개까지도 결합될 수 있으므로 명사 어형 생성 과정에서 과잉 생성의 문제가 발생한다. 생성 과정을 통제하기 위하여 KORMOR은 기본 조사에 결합 순서치(Order)를 할당하고 좌연접 요구(Requires) 조건을 명시하여 줌으로써 조사가 이미 도입된 좌측의 명사 어간과 적절히 결합될 수 있도록 장치되었다. 이 논문은 명사 어간과 조사의 이러한 결합 통제 조건을 논하고 KORMOR을 통해 명사 어형들이 어떻게 분석되고 생성되는가를 간단히 예를 보여 줌으로써 시스템의 충족성을 뒷받침한다.

**주요 어휘:** 어간, 접미사, 조사, 어형, 규칙 기반, 기본형, 복합형, 변이형, 이형태 생성 규칙, 어형 생성 규칙, 순서치, 좌연접 요구 조건, 제의 조건

## 1 규칙기반의 생성

이 논문은 국어의 명사 어간과 조사를 결합하여 정형(well-formed)의 어형(word form)들을 효율적으로 생성할 수 있는 규칙기반의 어형 생성 시스템이 어떻게 통제되어 운용되는가를 보여 주는 것이 그 목적이다.

동사의 어간(예: “걸”)은 반드시 어미(예: “는다”)와 결합해야만 어형이 될 수 있지만, 명사 어간(예: “사람”)은 조사(예: “은”)와 결합하지 않더라도 그 자체(예: “사람”)로 어형이 될 수 있다.

\*이 논문은 한국과학기술연구원 시스템공학연구소 국어공학센터의 1996년도 전문가 지원(연구책임자: 한국과학기술원 전산학과 최기선 교수)과 독일 Erlangen 대학 컴퓨터언어학과(주임: Roland Hausser 교수)의 재정 및 연구 지원을 받아 이루어진 연구 결과의 일부이다. 이 연구에 필요한 컴퓨터 환경은 독일 Tübingen 대학 전산 자료 처리 센터 ZDV의 은광희 선생의 많은 도움으로 설치되었고, 자료 조사는 한국과학기술원 전산학과 최용석 군과 국립국어연구원의 이준석 선생의 도움으로 이루어졌다. 고려대 언어학과 최재용 교수와 강범모 교수, 영어교육과의 정태구 교수, 그리고 언어학과 노승원 양이 이 논문을 읽고 평을 해 주었다. 연구 지원과 도움과 평을 해 주신 분들 모두에게 감사드린다.

(1) 동사:

- a. \*걷
- b. 걷는다

(2) 명사:

- a. 사람
- b. 사람은

따라서 명사 어간과 조사의 결합은 동사의 경우와는 달리 매우 간단한 것처럼 보인다.

그러나 국어는 교착어(agglutinative language)라서 명사 어간에 조사가 거듭 결합될 수 있기 때문에 그 결합 구조가 복잡하다.

(3) 명사 어간과 조사의 결합:

- a. 선생
- b. 선생-님-들-만-으로-는
- c. 선생-님-들-께서-부터-는-커녕
- d. 선생-님-들-께-로-부터-밖에-는-야

여기서 접미사인 “님”과 “들”을 빼더라도 명사 어간에 붙을 수 있는 조사는 적어도 6개가 된다.<sup>1</sup> 따라서 기본 조사가 원형과 변이형(예: “을”, “를”, “르”)을 포함하여 100개가 된다면 이론상으로 가능한 기본형과 복합형의 조사는 100<sup>6</sup>개가 된다. 그러나 부산대의 채영숙 등(1991)에 의하면 실제로 사용되는 조사의 총수는 1,938개밖에 되지 않는다. 이 수를 늘려 국어의 가능한 조사의 수를 넉넉히 4,000개로 잡더라도 100<sup>6</sup>개에서 4,000개를 뺀 나머지 결합형들은 제거되어야 하므로 명사 어간과 조사의 결합을 규칙기반의 생성 시스템으로 다루는 일은 쉽지 않다.

그러나 이 논문에서 명사 어형들의 생성을 위하여 규칙기반의 방법을 이용하려는 이유는 이러한 방법이 코퍼스에 기반을 둔 확률적 방법과 상호보완적으로 활용될 수 있기 때문이다. 예를 들면, 채영숙 등 (1991)의 조사 목록에는 “에게”로 시작하는 복합형은 114개, “게”로 시작하는 복합형은 129개, “로”로 시작하는 복합형은 83개, “으로”로 시작하는 복합형은 94개이다. 그러나 “게”는 “에게”의 축약형이고 “으로”는 “로”의 변이형이므로 이것들의 복합형 사이에 이러한 차이가 날 이유가 없다. 또 한 예를 들면, “에게서”로 시작하는 복합형은 69개이고, “한테서”로 시작하는 복합형은 “70”개로 한 개의 차이가 난다. 이러한 차이가 어디에서 생기는가를 검토해 보면, “한테서와”라는 복합형 조사는 코퍼스에서 채취된 반면에, “에게서와”라는 조사는 발견되지 않은 점을 지적할 수 있다. 규칙기반의 분석은 이러한 차이가 왜 생기는지를 쉽게 예측해 낼 수 있는 체계적인 척도가 된다.

<sup>1</sup>채영숙 등 (1991)의 연구에 의하면 서로 결합 가능한 조사의 수는 최대 6개가 된다.

따라서, 코퍼스에 의존한 확률적 분석은 규칙 기반의 분석에 의해 얻어진 언어의 규칙적 현상의 사실성을 확인하기 위한 목적 지향적 연구를 가능케 해 준다.

## 2 조사의 기본형

규칙 기반의 명사 생성 시스템을 구축하려면 먼저 조사의 기본형 목록이 작성되어야 한다. 채영숙 등 (1991)은 국어의 조사 1,938개를 제시하면서 기본 조사가 56개가 된다고 주장하였다. 그러나 1,938개의 조사 중 어느 것이 복합형 조사이며 기본형 조사인지는 밝히지 않았다. 한영균과 유동석 (1993)의 연구도 금성출판사 (1991)의 <국어대사전>에서 조사와 어미를 추출해 냈다. 그러나 이 둘을 함께 열거하였기 때문에 어떤 항목이 조사인지 어미인지 구별이 되지 않는 경우가 있으며, 기본형 조사와 복합형 조사의 구별은 전혀 되어 있지 않다. 국어연구원 이준석 (1996)에 의하면 금성출판사의 <국어대사전>에는 360개의 조사가 수록되어 있고, 한글학회의 <우리말 큰사전>에는 393개의 조사가 수록되어 있다. 여기서 옛것과 방언이나 구어체라고 생각되는 것을 제거하면 기본형 조사를 금성판에서는 108개를, 한글학회 사전에서는 103개를 추려낼 수 있다.

본 연구는 기존의 이러한 연구들을 참조하여 108개의 조사를 기본형으로 잡았다. 이들 중에는 주격 조사 “이”와 “가”, 대격 조사 “을”, “를”, “르”, 그리고 주제 복수 조사 “은”, “는”, “ㄴ” 등을 유형별로 함께 묶지 않고 각각 별개의 항목으로 열거하였다. “로”, “로서”, “로써”에 매개 모음 “으”가 붙어서 생기는 “으로”와 “으로써”와 “으로써”의 경우도 “로”, “로서”, “로써”와 함께 모두 다 기본형으로 다루었다.

기본형 중에는 매개 모음 “이”로 시작하는 조사가 많다. 이것들도 기본형 108개 중에 수록되지만 어형 생성 시스템 **LEX**의 기초 모듈이 되는 어휘부(k\_particle.lex)에는 매개 모음이 없는 전자의 것만이 수록된다.

### (4) 매개 모음 “이”로 시작하는 조사:

ㄴ들-인들 ㄴ죽-인죽 ㄴ죽순-인죽순 ㄹ랑-일랑  
 고-이고 나-이나 나마-이나마 든-이든 든지-이든지  
 라든-이라든 라서-이라서 라야-이라야 랑-이랑  
 며-이며 시-이시 야-이야 야-이야(호격) 야말로-이야말로  
 여-이여 요-이요

“이”로 시작하는 이들 변이형 조사들은 모든 이형태들을 생성해 주는 전처리 과정에 의해 생성된다

### 3 **꺠꺠꺠**의 전처리 과정과 변이형의 생성

**꺠꺠꺠**은 언어 처리의 효율성을 극대화하기 위하여 어형 생성의 과정을 두 단계로 나눈다. 첫단계는 어형의 기본 요소인 형태소의 자질들을 기계적으로 명시해 줌으로써 이형태(allomorph)들을 생성해 주는 전처리 과정이다. 이 전처리 과정에서 입력되는 파일은 기본 어휘 목록(korean.lex)이며 여기에 이형태 생성 규칙(파일명: korean.all)이 적용된다.<sup>2</sup>

**꺠꺠꺠**의 전처리 과정에서 구체적으로 이루어지는 작업은 두 가지다. 첫째, 기본 어휘 목록(korean.lex)에 수록된 어휘 항목에 기계적으로 부여할 수 있는 규칙적인 자질(feature)들을 할당하고 둘째, 어형 생성 과정에서 변형을 필요로 하거나 새로운 자질을 도입하는 일 없이 어형 생성 규칙을 적용할 수 있도록 이 규칙에 입력될 이형태(allomorph)들을 미리 생성해 놓는다.

이 전처리 과정을 명사 어간에 적용하면, 명사 어형 생성에서 명사 어간들은 접미사나 조사들보다 제일 먼저 도입되므로 그 순서치로 0의 값을 할당받게 된다. 그리고, 많은 조사들이 어간의 음절 구조에 따라 어간과의 결합 가능성이 결정되므로 어간의 음절 구조를 미리 분석하면서 그 음절 자질을 어간에 명시해 줄 필요가 있다. 예컨대, 주격 조사 “가”는 명사 어간이 개음절(Syllable: open)로 끝나는 어간(예: “교수”)에는 붙을 수 있으나 폐음절(Syllable: closed)로 끝나는 어간(예: “선생”)과는 결합할 수가 없다.<sup>3</sup> 따라서, 기본 어휘 “교수”와 “선생”은 어간이므로 둘 다 그 순서치로 Order: 0이라는 동일한 자질을 부여받으면서도 음절 자질에 대해서는 서로 다른 자질을 할당받게 된다. 즉, 명사 어간 “교수”는 Syllable: open이라는 자질을 할당받는 반면에, 명사 어간 “선생”은 Syllable: closed라는 자질을 할당받는다.

끝으로, 동사 어간들과는 달리 명사 어간은 조사 없이도 단독으로 어형이 될 수도 있고 조사와 결합하여 어형이 될 수도 있다. 따라서 명사의 형태(Form)는 완료(terminal)와 어간(stem)이라는 두 개의 값을 취하게 된다. **꺠꺠꺠**은 다중치(multi-segment value)라는 개념을 도입하여 명사 형태의 이중적 값을 terminal&stem의 다중치로 표시한다.

다음은 **꺠꺠꺠**의 기본 어휘 목록(korean.lex)을 구성하는 그 일부가 되는 명사 어휘 목록(k\_noun.lex)에 이형태 생성 규칙이 적용되어 생성된 새로운 명사 목록(k\_noun.lex.cat)의 일부이다.

<sup>2</sup> 확장자 .all은 allomorphy(이형태)를 다루는 파일임을 말한다.

<sup>3</sup> 음절 구조는 개음절(open), 폐음절(closed), 유음절(liquid) 3가지로 나뉜다. 개음절은 음절이 모음으로, 폐음절은 “ㄹ”을 제외한 자음으로, 유음절은 “ㄹ”로 끝나는 음절을 뜻한다. 국어의 경우, 이러한 음절 자질들은 어간의 마지막 음절의 구조에 의해 결정된다.

(5) 기본 명사 목록: k\_noun.lex

- a. [Phon: "교수", Class: noun,  
Sem: [Content: <"교수", "professor">]];
- b. [Phon: "선생", Class: noun,  
Sem: [Content: <"선생", "teacher">]];

(6) 생성된 명사 목록: k\_noun.lex.cat

- a. [Phon: "교수", Baseform: "교수", Class: noun,  
Order: 0, Form: terminal&stem, Syllable: open,  
Sem: [Content: <"교수", "professor">]]
- b. [Phon: "선생", Baseform: "선생", Class: noun,  
Order: 0, Form: terminal&stem, Syllable: closed,  
Sem: [Content: <"선생", "teacher">]]

이 두 목록을 비교하여 보면, 기본 명사 목록 (5)와는 달리 전처리 과정을 통해 생성된 명사 목록 (6)에는 순서(Order), 형태(Form), 음절(Syllable)에 대한 자질 값이 새로이 도입되어 표시되어 있다. 이 자질들은 이형태 생성 규칙 korean.all에 의해 자동적으로 도입된 것이다.

매개 모음 “이”로 시작하는 조사들도 전처리 과정에 의해 생성된다. 다음은 그 한 예이다.

(7) 기본 조사 목록: k\_particle.lex

[Phon: "라도", Class: particle, Order: <17>, Bridge: i\_bridged,  
Requires: [Class: noun, Form: stem&AF&CF&DF&QF, Syllable: open],  
Results: [Form: terminal], Sem: [DF: even]];

기본 조사 목록에는 “이”로 시작하는 조사가 포함되어 있지 않다. 그 대신, 예를 들면 “라도”와 같이 “이”로 시작하는 변이형 “이라도”를 허용하는 조사는 Bridge: i\_bridged라는 자질이 표시되어 있다. 이 자질 때문에 **이**은 그 전처리 과정을 통해 다음과 같이 “라도”와 그 변이형 “이라도”를 생성해 낸다.

(8) 생성된 조사 목록: k\_particle.lex.cat

- a. [Phon: "라도", Baseform: "라도", Class: particle, Order: <17>,  
Requires: [Class: noun, Form: stem&AF&CF&DF&QF,  
Syllable: open],

Results: [Form: terminal], Sem: [DF: even]]

b. [Phon: "이라도", Baseform: "라도", Class: particle, Order: <17>, Requires: [Class: noun, Form: stem&AF&CF&DF&QF, Syllable: closed&liquid], Results: [Form: terminal], Sem: [DF: even]]

위에 새로이 생성된 두 개의 조사 “라도”와 “이라도”를 비교해 보면, “라도”의 속성 Requires 안에는 음절 자질이 개음절(Syllable: open)로 표시되어 있는데 반하여, “이라도”에는 음절 자질이 폐유음절(Syllable: closed&liquid)로 표시되어 있다.<sup>4</sup> 이 폐유음절의 자질은 변이형 조사 “이라도”를 생성할 때 부여된 자질이다.

Requires라는 속성과 Results라는 속성은 복합 속성으로, 그것들의 값은 속성과 그 값으로 구성된 여러 자질들로 구성된다. 이에 대한 설명은 5.2절에서 어간과 조사의 결합 조건을 설명할 때 상세히 논하겠다.

## 4 어간과 접미사와 조사의 순서

명사 어형은 어간-접미사-조사의 순으로 이루어진다.

(9) 결합 순서: 어간-접미사-조사

- a. 선생
- b. 선생-님
- c. 선생-님-들
- d. 선생-님-들-께서
- e. 선생-님-들-께서-만-은

위의 예를 보면, 어간이 제일 먼저 오고 그 다음에 접미사 “님”, 접미사 “들”, 조사 “께서”와 “만”과 “은”이 일정한 순서를 지키며 연결된다. 이 순서가 안 지켜지면 비정형의 표현이 생긴다.<sup>5</sup>

(10) 비정형:

- a. \*님-선생
- b. \*선생-들-님
- c. \*선생-님-들-은-께서-만

<sup>4</sup> 폐유음절은 폐음절이거나 유음절인 음절을 뜻한다.

<sup>5</sup> 복수 접미사 “들”은 비교적 자유로워서 명사 어형의 제일 끝이나 부사 다음에도 올 수 있다. 예를 들면, (a) “선생-님-들-께서”, (b) “선생-님-께서-들”, (c) “\*선생-님-께서-들-만”, (d) “너희는 잘 돌 있거라” 등이 있다.

어간과 접미사와 조사의 이러한 순서를 어형 생성 시스템  $\mathcal{G}$ 에 구현하기 위하여, 변이형들의 전처리 과정(korean.all)에서 체언과 용언의 모든 어간은 그 결합 순서치로 0의 값을 자동으로 할당받는다. 접미사의 순서치는 자동으로 할당되지 않고 기본 어휘 목록의 일부가 되는 접미사의 목록(k.suffix.lex)을 작성할 때 하나하나 수작업에 의해 명시된다. 따라서 기본 조사의 수가 적을수록 작업하기가 수월하다.

어간의 순서치는 0이므로 접미사의 순서치 영역은 0보다 큰 1과 9 사이로 넉넉히 잡아 두면 된다.

(11) 어간과 접미사의 연결 순서:

| stem        | suffixx | suffix | suffix |
|-------------|---------|--------|--------|
| 순서치 Order 0 | 1       | 2      | 3      |
| 선생          | 님       |        | 들      |
| 홍길동         | 이       |        |        |
| 도둑          |         | 놈      | 들      |
| 용팔          | 이       | 놈      | 들      |

여기서 “님”은 존대 접미사이고 “놈”은 하대 접미사이다. “홍길동이가”의 “이”는 주격 조사와는 다른 일종의 멸시 접미사로 우리말식 이름 뒤에 붙는다.<sup>6</sup> “들”은 복수 접미사로 비교적 뒤에 붙으므로 일단 3의 순서치를 배정해 놓았다.<sup>7</sup>

채영숙 등 (1991)에 의하면 조사는 6개까지 연결될 수 있으므로 그 순서치 영역에는 적어도 6개의 자리가 필요하다. 그리고 조사는 접미사 다음에 오므로 조사의 순서치 영역을 10에서부터 19까지 10개의 자리를 여유있게 잡아 놓았다.

특수 조사 “만”은 그 위치가 복잡하다. 앞의 예에서는 “만”이 주격 조사 “께서” 다음에 왔으나 일반적으로 격조사 앞에 온다.

(12) “만”의 위치:

- a. 선생-님-께서-만
- b. 선생-님-께-만
- c. 선생-님-만-이
- d. 선생-님-만-을
- e. 선생-님-만-의
- f. 선생-님-만-은

<sup>6</sup>외국 이름이라도 “모택동이가”처럼 우리말식으로 불리면 멸시 접미사 “이”를 붙일 수 있으나 “클린톤”과 같은 외래명에 붙이면 좀 어색하게 들린다. 예를 들면, “클린톤이가 사담이를 때려 부셔라”는 말은 어색하다.

<sup>7</sup>“들”이 부사 뒤(예: “잘들”)나 어미 뒤(예: 먹어들 보게!)에 붙는 경우는 다른 논문에서 차후에 다루겠다.

위의 예를 보면, 조사 “만”이 “께서”와 “께”의 경우는 그 다음에 오고, 격조사 “이”, “을”, “의”와 특수조사 “은”의 경우에는 그 앞에 온다.

특이하게도, 도구 조사 “으로”의 경우는 조사 “만”이 그 앞에 올 수도 있고 뒤에 올 수도 있다.<sup>8</sup>

(13) 도구 조사 “으로”와 “만”의 결합:

- a. 물 없이 빵만으로는 살 수 없다.
- b. 돈이 없어서 빵으로만 배를 채웠다.

따라서 “만”의 순서치는 “으로”의 순서치보다 작아야 하기도 하고 커야 하기도 한다.

채영숙 등 (1991)의 자료를 근거로 하여 기본 조사들의 순서치를 분석한 결과(부록 1 참조)의 일부를 보이면 다음과 같다.

| No | 10 | 11  | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|----|----|-----|----|----|----|----|----|----|----|----|
| 1  |    |     |    |    |    |    |    | 이  |    |    |
| 2  |    |     |    |    |    |    |    | 가  |    |    |
| 3  |    | 께서  |    |    |    |    |    |    |    |    |
| 4  |    | 께옵서 |    |    |    |    |    |    |    |    |
| 5  | 께  |     |    |    |    |    |    |    |    |    |
| 6  |    |     |    |    |    |    |    | 을  |    |    |
| 7  |    |     |    |    |    |    |    | 의  |    |    |
| 8  |    |     |    |    |    |    |    | 은  |    |    |
| 9  |    |     |    | 으로 |    |    |    |    |    |    |
| 10 |    |     | 만  |    |    |    | 만  |    |    |    |

이 표에 의하면 “만”은 그 순서치가 12와 16이 된다. “선생님께서만”과 “선생님만이”의 차이를 다루기 위해서는 “만”의 순서치가 “께서”의 순서치 11과 17 사이에만 있으면 된다. 그러나 “으로”의 경우에는 “만”이 앞이나 뒤에도 올 수 있으므로 순서치가 둘이 필요하다. 그리고, “만”의 둘째 순서치가 16이 되는 이유는 “빵으로만은”과 같은 경우가 있어서 “만”의 순서치가 “은”의 순서치 17보다 앞에 있어야 하기 때문이다.

## 5 Malaga와 결합 조건의 명시

☐☐☐은 한국어의 어형 분석-생성 시스템으로 자연 언어의 처리에 적합하도록 C-언어를 보강한 Malaga라는 프로그래밍 언어로 구현되었다. 이 언어는 첫째, 자연 언

<sup>8</sup> 조사 “으로”는 도구의 뜻 이외에 방향 조사로 쓰일 수도 있고 결정 조사로 쓰일 수도 있다. 예를 들면, “사람들이 서울로 몰려 든다”(방향), “김길동을 대표로 뽑았다”(결정) 등인데 이 경우는 “만”이 “으로” 다음에만 붙는다. 따라서, “사람들이 서울로만 몰려 든다”는 정문이지만 “사람들이 서울만으로 몰려 든다”는 비문이다.



어 표현들의 자질 구조(feature structure)를 간결하게 기술할 수 있는 여러 유형의 변항과 상항, 그리고 자질 구조에 대한 여러 가지 연산들을 C-언어로 정의하여 준다. 따라서 컴퓨터는 Malaga 언어로 표현된 어간과 조사의 자질 구조를 인식할 수 있으며 자질 구조에 대한 연산들을 수행할 수 있다.

## 5.1 순서 조건

그 두드러진 예가 결합 순서와 관련된 연산이다.

### (14) 결합 순서 조건

```
foreach $Position in $Right.Order:
  if $Left.Order less $Position then
    set $Order := $Position;
  end;
end;
```

이 규칙을 보면, \$로 표시된 변항들 이외에, “if...then”과 같은 볼리언 연산뿐 아니라 전칭 양화사 “foreach”와 집합론에서의 원소 관계를 나타내는 “in”이 이미 정의되어 쓰이고 있다. 그밖에도 수의 크기를 비교할 수 있는 “=”(같음)와 “less”(더 작음)라는 연산들이 또한 정의되어 쓰이고 있다. 특히 후자의 연산 때문에 어간과 접미사와 조사의 결합 순서가 결합 조건으로 제시된다.

여기서 단순히 “? \$Left.Order less \$Right.Order”라 묻지 않고 앞에서와 같이 순서에 관한 제약이 복잡해 진 이유는 “만”과 같은 조사는 순서치가 하나가 아니고 둘, 즉 <12,16>이기 때문이다. 가령 순서치 13의 어간 “뿡으로”가 “만”과 결합하려면 “만”의 순서치 중에는 16이 있으므로 순서 제약 조건이 충족되고 새로 결합된 “뿡으로만”의 순서치는 16이 된다. 그리고 이 순서치가 0이 아니므로 다음 단계로 넘어간다.

## 5.2 좌연접 요구(Requires) 조건

Malaga의 두번째 특징은 범주 문법을 자연언어 처리에 적합하도록 재형식화한 Hausser (1992)의 좌연접 문법을 그 이론적 기반으로 삼고 있다. 따라서 이 언어는 범주 문법의 기본 결합 연산인 범주 삭제(cancellation) 조건을 쉽게 표현할 수 있는 강력한 기술력(descriptive power)을 갖고 있다.

좀 더 구체적으로 설명하면, 범주 문법에서 범주 A의 표현  $\alpha$ 가 어떤 표현  $\beta$ 와 결합하려면 이  $\beta$ 는 B/A의 함수 표현이라야 한다. 일반적으로, 범주 문법의 삭제 규칙은  $(A, B/A) \Rightarrow B$ 의 형식을 취하며 전자를 논항 범주, 후자를 함수 범주라고

한다. 이러한 형식을 좌연접 문법을 기초로 한 **ㅁㅁ**에 적용하면 어간은 논항 범주가 되고 접미사나 조사는 함수 범주가 된다.

접미사나 조사의 함수성은 두 개의 복합 속성 Requires와 Results에 의해 표현된다. 복합 속성 Requires는 함수 범주 B/A의 A에 해당하고 복합 속성 Results는 B에 해당한다. 따라서, **ㅁㅁ**에서 어간이 조사와 결합하려면 조사의 복합 속성 Requires에 들어 있는 모든 자질들이 어간의 자질들과 일치(match)해야 한다. 만일 이것이 성공하면, 그 결과 어간과 조사가 결합되며 새로이 결합된 표현은 복합 속성 Results에 들어 있는 자질들을 물려 받게 된다.

### 5.3 제외 조건 Restricts

명사 어간과 접미사 또는 조사의 결합은 그 순처치나 복합 속성 Requires의 자질에 의해서만 제약되지 않고 Restricts라는 조건에 의해서도 제약된다. 이 조건은 주어진 접미사나 조사를 결합 대상에서 제외시킨다. 예를 들면, “선생님”과 같이 이미 존재 접미사 “님”이 붙어서 존재 대상(Honored: referent)의 표현으로 분석되었을 경우에는 “선생님-놈”과 같은 잘못된 표현이 생성되지 않도록 접미사 “놈”에 제외 조건이 명시된다. 즉, “놈”의 자질 속에 Restricts: [Honored: referent]라는 복합 자질이 명시된다.

또 한 가지 예로, “만”이 잘못 거듭되는 “빵만으로만”의 경우를 들겠다. 순서치나 좌연접 요구 조건만으로는 위와 같은 비정형의 표현을 막을 수 없다. 그러나 “만”의 자질 구조 속에는 Restricts: [QF: only]라는 제외 조건이 들어 있다. 그리고 “빵만으로”라는 표현의 의미 자질에도 이미 [QF: only]라는 자질이 들어 있으므로 “빵만으로”에 “만”이 다시 결합될 수 없게 된다.

## 6 어형의 분석과 생성: 그 예

**ㅁㅁ**의 어형 생성 모듈인 korean.mor에는 결합 순서 조건뿐 아니라 복합 속성 Requires에 의한 좌연접 요구 조건, 복합 속성 Restricts에 의한 제외 조건들이 Malaga 언어가 이해할 수 있는 규칙의 형식으로 명시되어 있다.<sup>9</sup> 이 모듈은 하나의 시스템으로 어형의 분석과 생성이 가능하다. 형태소 분석을 하라는 명령 ma(morphological analysis)와 함께 어형이 입력되면 이 어형이 명사일 경우에 어간과 접미사, 조사 등으로 분석하여 그 결합 과정과 자질 구조를 보여 준다. 가능한 어형을 생성하려고 할 경우에는 paradigm이라는 명령과 함께 어간을 입력시켜 주면 그 뒤에 붙을 수 있는 요소들을 골라 가면서 가능한 모든 결합형을 생성해 낸다.

<sup>9</sup>korean.mor의 확장 부호 .mor은 morphology를 줄인 말이다.

## 6.1 명사 어형의 분석

다음은 Malaga를 작동시켜 몇몇 명사 어형을 분석해 본 결과이다.

(15) “선생님만이”의 분석:

```
malaga> ma 선생님만이
malaga> result
result(s) for "선생님만이":
1: "선생/님/만/이",
[Phon: "선생님만이", Baseform: "선생", Class: noun,
GF: subject, Cas: nominative, Sem: [QF: only,
Honored: referent, Content: <"선생", "teacher">]]
```

malaga에 ma 선생님만이이라고 입력을 했더니 위와 같은 결과가 나왔다. 첫째, 입력된 “선생님만이”가 “선생/님/만/이”로 분석되었다. 그 자질 구조를 보면, 이 어형은 주격(nominative) 명사(noun)로서 주어(subject)의 문법적 기능을 가졌으며 지칭되는(referent) “선생”(teacher)이 존대받는(Honored) 유일한(only) 대상으로 해석된다.

위의 “선생님만이”의 경우와는 달리, 주격 조사가 “께서”인 경우에는 조사 “만”이 그 다음에 온다. 따라서, “선생님께서만”은 아래에서 보여 주듯이 제대로 분석되었다.

(16) “선생님께서만”의 분석:

```
malaga> ma 선생님께서만
malaga> result
result(s) for "선생님께서만":
1: "선생/님/께서/만", [Phon: "선생님께서만", Baseform: "선생",
Class: noun, GF: subject, Cas: nominative,
Sem: [ArgRole: source, QF: only, Person: yes,
Honored: referent, Content: <"선생", "teacher">]]
malaga>
```

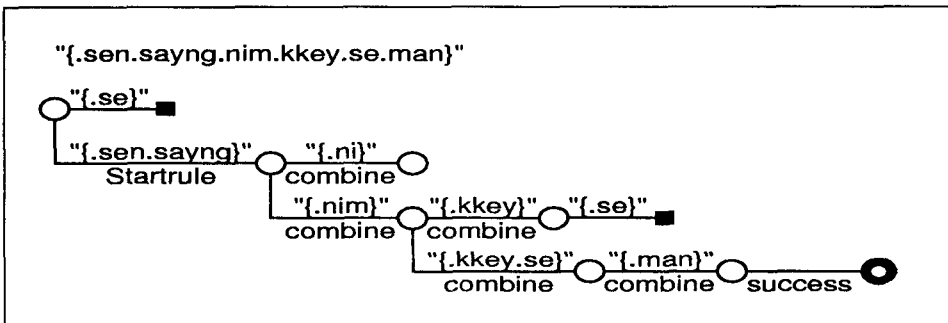
그러나 조사 “만”이 “께서”보다 앞설 수는 없다. 따라서 “선생님만께서”의 분석은 실패했다.

(17) “선생님만께서”의 분석:

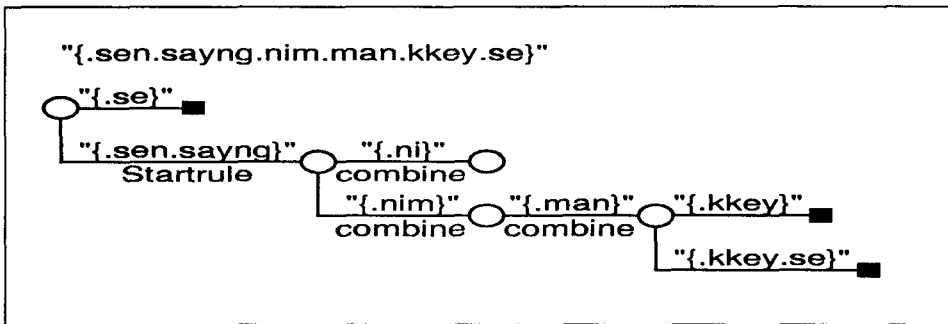
```
malaga> ma 선생님만께서
malaga> result
no result for "선생님만께서"
malaga>
```

그 실패한 원인을 알려면, 분석된 과정들을 비교해 보면 된다.

(18) “선생님께서만”의 수형 분석



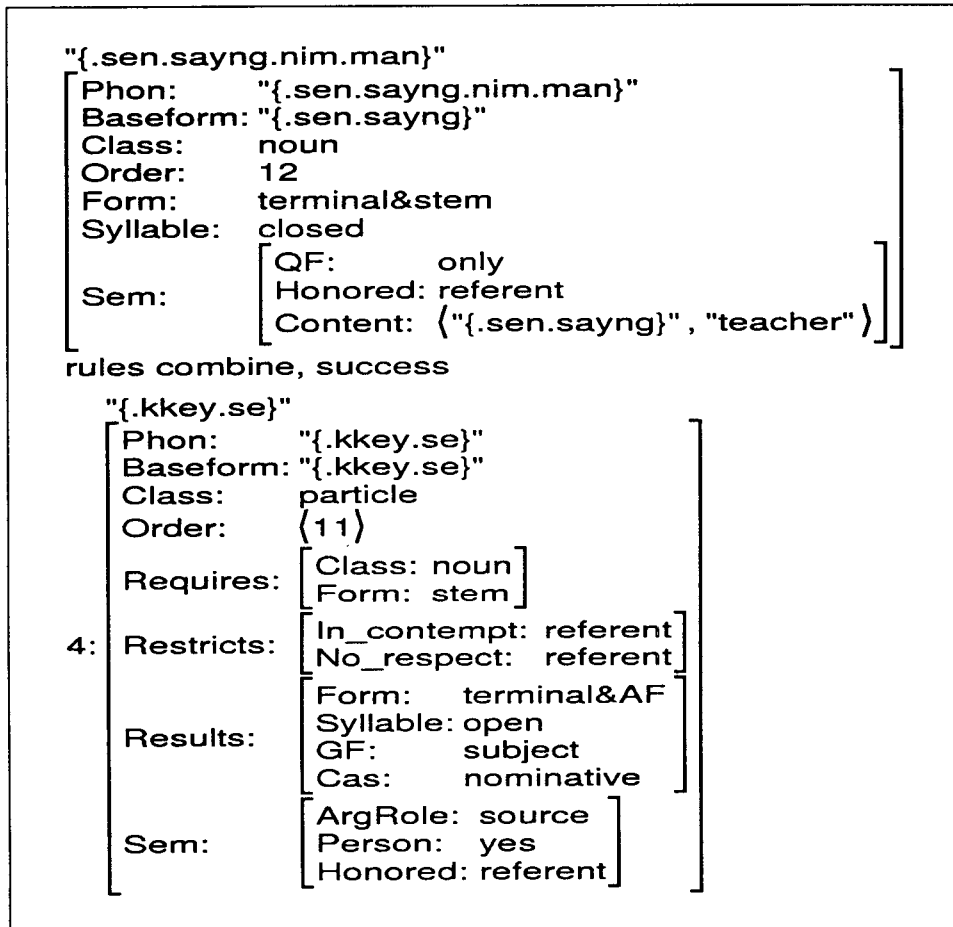
(19) “선생님만께서”의 수형 분석



위의 수형 분석에서 앞의 것은 두 개의 원이 겹친 쌍원 표시로 끝난 가지가 있다. 쌍원으로 끝난 이 가지는 success라는 말이 말해 주듯이, 입력된 “선생님께서만”이 “선생-님-께서-만”으로 성공적으로 분석되었음을 말해 준다. 반면에, 뒤의 수형 분석은 입력된 “선생님만께서”가 “선생-님-만”까지는 성공리에 결합 규칙(combine)이 적용되어 제대로 분석되었으나 여기에 “께서”를 결합시키는 과정은 실패했음을 말해 준다.

그 실패 요인을 좀 더 자세히 알아 보려면 다음의 자질 구조 분석도를 검토하면 된다.

(20) “선생님께서만”의 자질 구조도



위에서 좌연접 요구 조건은 충족되었다. 그러나 “선생님만”의 순서치는 12인데 “께서”의 순서치는 11이므로 이들의 결합은 순서 제약 조건에 걸려서 실패했다.

## 6.2 명사 어형의 생성

Malaga는 하나의 시스템으로 언어 분석과 생성이 가능하다. 따라서 한국어의 **ㄱ**도 어형 분석뿐 아니라 생성이 가능하다. 구체적으로 **ㄱ**에서 malaga를 작동

시켜 paradigm 선생이라는 명령을 주면 어형 생성 규칙의 generate라는 연산이 거기에 등재된 접미사나 조사 또는 어미를 보고 명사 “선생”과 결합이 가능한 어형들을 생성해 낸다. 다음은 실험을 위하여 접미사를 “님”과 “들” 두 개로 제한하고, 조사를 호격 조사 “여”와 이와 관련된 존칭사 “시”와 그 변이형 “이시”만을 생성 대상에 등재했을 때 얻은 결과이다

(21) 명사 어간 “선생”에서 생성된 어형

malaga> paradigm 선생

- 1: "선생",  
[Phon: "선생", Baseform: "선생", Class: noun,  
Sem: [Content: <"선생", "teacher">]]
- 2: "선생/님",  
[Phon: "선생님", Baseform: "선생", Class: noun,  
Sem: [Honored: referent, Content: <"선생", "teacher">]]
- 3: "선생/님/들",  
[Phon: "선생님들", Baseform: "선생", Class: noun,  
Sem: [Number: plural, Honored: referent,  
Content: <"선생", "teacher">]]
- 4: "선생/님/들/이여",  
[Phon: "선생님들이여", Baseform: "선생",  
Class: noun, GF: vocative, Sem: [Number: plural,  
Honored: referent, Content: <"선생", "teacher">]]
- 5: "선생/님/들/이시/여",  
[Phon: "선생님들이시여", Baseform: "선생", Class: noun,  
GF: vocative, Sem: [Number: plural, Honored: referent,  
Content: <"선생", "teacher">]]
- 6: "선생/님/이여",  
[Phon: "선생님이여", Baseform: "선생", Class: noun,  
GF: vocative, Sem: [Honored: referent,  
Content: <"선생", "teacher">]]
- 7: "선생/님/이시/여",  
[Phon: "선생님이시여", Baseform: "선생", Class: noun,  
GF: vocative, Sem: [Honored: referent,  
Content: <"선생", "teacher">]]
- 8: "선생/들",  
[Phon: "선생들", Baseform: "선생", Class: noun,  
Sem: [Number: plural, Content: <"선생", "teacher">]]

- 9: "선생/들/이여",  
 [Phon: "선생들이여", Baseform: "선생", Class: noun,  
 GF: vocative, Sem: [Number: plural,  
 Content: <"선생", "teacher">]]
- 10: "선생/들/이시/여", [Phon: "선생들이시여", Baseform: "선생",  
 Class: noun, GF: vocative, Sem: [Number: plural,  
 Honored: referent, Content: <"선생", "teacher">]]
- 11: "선생/이여",  
 [Phon: "선생이여", Baseform: "선생", Class: noun,  
 GF: vocative, Sem: [Content: <"선생", "teacher">]]
- 12: "선생/이시/여",  
 [Phon: "선생이시여", Baseform: "선생", Class: noun,  
 GF: vocative, Sem: [Honored: referent,  
 Content: <"선생", "teacher">]]
- malaga>

## 7 맺음말

본 시스템은 Hausser (1989)의 좌연접 문법에 기초한 Beutel 등(1996)의 Malaga Version 2.0으로 구현되었다. 이 Malaga의 원조는 Schüller 등 (1995)의 좌연접 파서 LAP이다. 그동안에 발표된 이기용의 논문들(1994, 1995a,b)은 규칙적 표현(regular expression)들의 일치(matching) 연산을 바탕으로 한 LAP에 의존한 연구 결과였으며 그 후의 논문들(이기용 1995c, 1995d)은 Malaga에 의존한 것들이다.

LAP과 Malaga의 근본적인 차이는 전자가 규칙적 표현들만을 분석 대상으로 받아들일 수 있고 그것들 사이의 연산밖에 다룰 수 없는 반면에, 후자는 속성과 그 값으로 구성된 자질과 자질 구조들을 분석 대상으로 수용할 수 있고 이것들 사이의 연산 처리가 가능하다는 점이다. 더욱이 최신판 Malaga Version 2.0은 결합 순서와 관련된 연산 "less"(더 작다)를 정의해 주고 있으며, 자질들을 그 값으로 취할 수 있는 "Requires"나 "Results"와 같은 복합 속성을 정의해 줄 수 있다. 따라서, 이 논문에서 간단히 보여 준 바와 같이 언어 표현들의 결합을 효율적으로 통제해 줄 수 있는 제약 조건을 명시하기가 쉬워졌다.

현재의 시스템에는 기본형 조사 108가 수록되어 있다. 이것들을 생성(generate) 연산의 결합 대상 목록에 등재했을 때 실험 결과로는 3,000여개의 복합형 조사들이 나왔다. 이것들 중 과연 어떤 것들이 실제로 쓰이는가를 검토하기 위해선 한국어의 커다란 데이터베이스를 기초로 하여 얻은 통계적 자료와의 비교 연구가 뒤

따라야 할 것이다. 다행히 Malaga에는 커다란 텍스트를 분석하여 확률적 통계를 제시해 주는 기능도 있다. 따라서 규칙 기반의 **Malaga**은 이러한 통계적 처리와 접목됨으로써 응용 가능한 한국어 처리 시스템이 되리라 기대된다.

## 참고 문헌

1. 이기용 (1994), HLAP: Left-Associative Parser for Korean in Hangul, 서울대학교 어학연구소 제28회 어학연구회.
2. 이기용, 박종원 (1995a), 한국어 어형 생성시의 과잉 생성의 문제, HCI '95 학술 대회 발표 논문집, 42-54.
3. 이기용 (1995b), Hausser의 좌연접 방식에 의한 한국어 용언 축약형의 생성, 1995년도 인지과학회 춘계 학술 발표 논문집, 239-247.
4. 이기용 (1995c), Malaga에 의한 이형태와 어형의 생성, 국어학회 제22회 국어학 공동 연구회 발표 요지, 39-54.
5. 이기용 (1995d), Lee (1995)를 볼 것.
6. 채영숙, 김재원, 김민정, 김지영 (1991), "한국어 조사 유형 사례집," PNU AI TR91-1.
7. 한영균, 유동석 (1993), 컴퓨터를 이용한 국어 자료 분석에 대한 기초적 연구, 서울: 국립국어연구원.
8. Beutel, B., G. Schüller, and R. Hausser (1996), "Malaga Version 2.0: An Implementation Language for Left-Associative Grammar"(unpublished), Erlangen: Universität Erlangen-Nürnberg, Abteilung Computerlinguistik.
9. Hausser, R. (1989). *Computation of Language*, Berlin: Springer-Verlag.
10. Lee, Kiyong (1995). "Recursion Problems in Concatenation: A Case of Korean Morphology," *Language, Information and Computation: Proceedings of The 10th Pacific Asia Conference in Hong Kong*, 215-224.
11. G. Schüller and R. Hausser (1995), "LAP: Left-Associative Parser," Erlangen: Universität Erlangen-Nürnberg, Abteilung Computerlinguistik.