

낱말·패러다임 형태론에서의 굴절 부류 설정

노 용균

서울대학교 언어학과

Deciding on Inflectional Classes in a Word-and-Paradigm Morphology

Yongkyoon No

Department of Linguistics, Seoul National University

요약

한국어 동사의 굴절은 (1) 접사의 누적 (2) 어간과 접사의 변이형 (allomorphy) (3) 원격한 접사의 끝이 어간에 민감한 점 등의 이유로 단순한 접사 추출 (affix-stripping)에 의한 형태소 분석이 감당하기 어려운 부분을 갖고 있다. 따라서 형태소에 바탕을 둔 형태이론보다 더 세련된, 낱말에 바탕을 둔 형태이론이 요구 되는데, 이 논문에서는 “낱말·패러다임 형태이론 (Word and Paradigm Morphology)”에서 실현규칙(rules of realization)이 언급하는 요소 중의 하나인 굴곡 부류(inflectional classes)가 어떻게 설정 되어야 할 것인가를 다룬다. 특정한 형태·통사적 속성들(morpho-syntactic properties)이 실현될 때 (1) 주요 어간(primary stem)이 요구되느냐 아니면 부차 어간(secondary stem)이 요구되느냐, 그리고 (2) 접사의 어떤 끝이 요구되느냐에 바탕을 두고 한국어의 동사를 열 다섯 개의 굴절 부류로 분류한다.

I. 서론

형태소에 바탕을 둔 형태 이론 (morpheme-based morphology)은 의미와 음운 사이의 단순한 대응 관계를 상정하는 만큼 모델의 복잡도가 낮아서 이론언어학에서건 전산언어학에서건 널리 채택되고 있다. 그러나 자연언어는 이러한 모델이 감당할 수 없는 다음과 같은 몇 가지 현상을 갖는다. (괄호 안의 예는 우리 말에서 끌어 왔다.)

1. 한 어휘소 (lexeme)가 두 개 이상의 어간 (stems)을 갖는다.
(멀 `be far' -- /멀/과 /머/로 나타남)
2. 한 접사가 두 개 이상의 끝으로 나타난다.
(감탄형 -- /는군/과 /군/으로 나타남)
3. 접사의 끝이 선택될 때 인접하지 않은 어간에 좌우된다.
(/잡는구나/, /잡으시는구나/, /잡겠구나/, */잡으시구나/, */잡겠는구나/)

4. 임의의 굴절꼴이 없는 낱말이 있다.
(/데리고/, /데려/, */데린다/, */데립시다/)
5. 한 무리의 형태통사적 속성들이 누적적으로 실현된다.
(주체 존대, 명령형 /으십시오/: /서십시오/, */십시오/)

이와 같은 현상들을 수용하는 형태 이론은 낱말에 바탕을 둔 (word-based) 것이어야 하는데, [3], [4], [5], [7] 등의 “낱말·패러다임 형태 이론 (Word and Paradigm Morphology)”이 유망해 보인다.

낱말·패러다임 형태이론에서는 실현규칙 (rules of realization)을 써서 형태 동사적 속성들 (morpho-syntactic properties)을 음운연쇄체에 연결시키는데, 실현규칙이 언급하는 항목은 (i)형태·통사적 속성, (ii)굴절의 상태(state of inflection), (iii)낱말의 굴절 부류 (inflectional class), (iv)지정된 어간, 그리고 (v)음운적 조작 (phonological operations)이다. 굴절의 상태란 낱말의 굴절이 완료되었는지 아니면 다른 실현규칙들이 추가적으로 적용되어야 하는지에 관한 정보로서 굴절을 유한상태 자동자 (Finite State Automaton)로 표시할 때의 ‘상태’와 같다. 음운적 조작으로는 점두사나 접미사를 첨가하는 것을 비롯해서, 모음 교체, 음소 삭제, 중첩, 무조작 (identity mapping) 등이 있다. 지정된 어간이 언급되는 이유는 위의 (1)에서 보았듯이 한 어휘소가 두 개 이상의 어간을 갖는 경우가 있기 때문이다.

형태부문의 실현규칙은 따라서 아래와 같은 모습을 갖게 된다.

해당 낱말이 (a) 형태·통사적 속성들의 집합 μ 를 갖고, (b)굴절 부류들의 집합 L 의 어떤 구성원에 속하면, σ_1 로부터 ω 에 의하여 σ_j 가 도출된다. 이 때 요구되는 어간은 s 이다. [σ_1 와 σ_j 는 굴절의 상태들에서 값을 취하고 ω 는 음운적 조작들에서 그 값을 취한다.]

두 개 이상의 어간 중에서 어느 것이 선택되는지조차 명시되어야 할 경우는 음운적 조작 ω 가 어간에 적용되는 접사 첨가의 조작일 경우와 무조작 (identity mapping)인 경우다. 특정한 실현 규칙이 어간을 지정함으로써 과다 생성 (overgeneration)이나 과다 수용 (overrecognition)을 막을 수 있는 것이다. 동사 살 ‘live’로부터 얻은 *살습니다/, *살니다/, */사습니다/를 다 배제하고 오로지 /삽니다/만을 생성·수용하는 것이 그것이다.

실현 규칙의 틀이 위와 같이 정해졌다고 해도 한 언어의 굴절 부문을 완전히 기술하려면 몇 가지의 구체적인 문제를 해결하고 모든 실현 규칙을 서술해야 한다. 해결해야 될 문제들로는 (i)모든 음운 현상을 어휘적 형태적 제한이 있는 형태음운 과정 (morphophonemic processes)과 오직 음운적 구역(phonological domains)의 영향만 받는 자동음운 과정 (autophonological processes)으로 양분하는 것과 (ii)이를 바탕으로 모든 어휘소를 하나의 굴절 부류에 할당해 주는 것이 있다. 이 글에서는 /으/탈락 (또는 삽입), 자음

군 간략화(consonant cluster simplification), /ð/의 /w/로의 약화, /r/탈락, 반모음화 등의 음운 과정을 형태 음운 과정으로 보고, 구개 저해음 중화, 비음 앞 저해음의 비음화, 추상적 음소 /#/ 또는 /h/이 뒤따르는 저해음과 합류하여 각각 경음 및 유기음을 낳는 등의 과정을 자동적 과정으로 보고 한국어 동사의 굴절 부류를 설정하고자 한다.

II. 굴절 부류 설정의 기준

동일한 굴절 부류에 속하는 모든 낱말은 굴절 패러다임의 어떤 항에서도 접사의 동일한 변이형(allomorph)과 동일 순위의 어간을 사용한다. 변이형이 하나밖에 없는 접사는 굴절 부류의 설정에 아무런 역할을 하지 않는다. 변이형이 둘 이상인 접사는 어느 변이형이 어떤 무리의 어휘소에 쓰이느냐에 따라 어휘소를 분류할 기준을 제공한다. 구체적인 예를 들면, 우리 말의 현재 관형형 접사는 /는/이나 /은/ 또는 /ㄴ/으로 나타난다. 모든 동작동사와 존재사는 /는/과 결합하고 상태동사(혹자는 형용사라고도 함)와 지정사는 /은/ 또는 /ㄴ/을 요구한다. 뒤 두 변이형 (/은/과 /ㄴ/)의 선택은 어기(base)의 마지막 소리가 자음이냐 모음이냐에 따라 체계적으로 결정되므로 해당 실현 규칙의 음운조작 부분을 운용하여 기술할 수 있지만, /는/이 선택되는지 아니면 /은/이나 /ㄴ/이 선택되는지에 관한 정보는 어휘소마다에 하나하나 주어야 한다.

변이형이 둘 이상인 접사가 모두 굴절부류의 설정에 한 몫을 하는 것은 아니다. 예컨대, 우리말의 대조 현재 서술 연결형 접사는 /는데/나 /은데/ 또는 /ㄴ 데/로 나타나는데, 이 변이형들은 동사를 분할하는 방식이 방금 살펴 본 현재 관형형 접사가 분할하는 방식과 전적으로 동일하다.

어간의 끝이 둘 이상인 어휘소도 굴절 부류의 설정에 중요한 역할을 한다. 주요 어간(primary stem)과 부차어간(secondary stem)은 상보적 분포를 보이는데, 어느 어간이 어떤 형태동사적 특성들을 실현하는 데에 쓰이느냐에 따라 어휘소를 분류할 수 있는 것이다. 예를 들어 우리 말의 동사 하 'do'와 /하/를 마지막 음절로 갖는 모든 동사(옴하 'come close to', 삼하 'decay' 등)는 /하/와 /해/ (/__하/와 /__해/)를 변이형으로 갖는다. 동사 굴절 패러다임 전체를 놓고 볼 때 부차 어간 /하/가 선택되는 곳은 (i)일부 보조 동사가 지배하는 꼴 (해 보고), (ii)반말체 종결형, (iii)이유 연결형 (/어서/, /아서/, /서/), (iv)양보형 (/어도/, /아도/, /도/), (v) 과거형 (/었/, /았/, /ㅁ/), (vi) 하대 명령형 (/어라/, /아라/, /라/), (vii) 감탄형 (/어라/, /아라/, /라/), 그리고 (viii) 대과거형 (/있었/, /았었/, /ㅁ었/)에서이다. 이 밖의 모든 곳에서 주요 어간 /하/가 선택된다. 두 개의 어간이 쓰인다는 점에서 하와 동일한 동사 굶 'skip a meal'의 변이형별 분포는 이와 다르다.¹⁾ 이 어휘소의 부차 어간 /굶/은 위의 (i) - (iv)에서만 선택되는 것이 아니라 (굶어

1. 굶의 굴절 꼴들은 [굶지만], [굶므면], [굶므시], [굶머도], [굶눔군], 그리고 [굶는데]로 발음되는데, 첫번 째 꼴 [굶지만]을 얻기 위해 주요 어간을 /굶#/로 설정하고, 추상적인 음소 /#/가 뒤따르는 저해음을 경음화하는 자동 음운 과정을 아울러 설정한다. 추상적인 이 음소는, 마치 /h/이 그렇듯이, /ㄴ/ 앞에서 삭제된다고 본다.

버리고, 끓어, 끓어서, 끓어도), 조건 연결형에서와 (끓으면, *끓으면) 객체 존대형 (끓으시, *끓으시)에서도 쓰인다. 동사 하의 부차 어간 (/해/)은 이 자리들에서 쓰이지 않음에 유의하면(*해면, *해시), 주요 어간과 부차 어간의 분포가 어떤가에 따라 동사들이 분류될 수있음을 볼 수 있다.

III. 결과 및 논의

앞 절에서 제시한 기준에 의해 우리 말의 동사를 분류하면 아래의 굴절 부류들을 얻는다. 이 표에서 숫자 1, 2, 3은 각각 주요 어간, 제 2어간, 제 3어간을 일컫는다. 접사의 끝이 어간의 음운정보로부터 예측할 수 없는 경우에는 제시되었다.²⁾

굴절 부류	주요 어간	부차 어간	지만	으면/면	으시/시	어도/아도/도/라도	는군/군	는데/은데/ㄴ데	어라/아라/라/거라/너라
I	잡		1	1	1	1	1 는군	1 는데	1
II	잡		1	1	1	1	1 군	1 은데	1
III	이		1	1	1	1 라도	1 군	1 ㄴ데	1
IV	있		1	1	1	1	1 군	1 는데	1
V	업	없	1	2	2	2	1 군	1 는데	2
VI	가		1	1	1	1	1 는군	1 는데	1 거라
VII	오	와	1	1	1	2	1 는군	1 는데	1 너라
VIII	부르	불리	1	1	1	2	1 는군	1 는데	2
IX	빠르	빨리	1	1	1	2	1 군	1 ㄴ데	2
X	놀	노	1	1	2	1	2 는군	2 는데	1
XI	멀	머	1	1	2	1	1 군	2 ㄴ데	1
XII	끓#	끓	1	2	2	2	1 는군	1 는데	2
XIII	점#	점	1	2	2	2	2 군	2 은데	2
XIV	걸	걸으, 걸어	1	2	2	3	1 는군	1 는데	3
XV	밧	미우, 미워	1	2	2	3	1 군	2 ㄴ데	3

2. 양보 연결형 /어도/, /아도/, /도/는 /라도/에 대응한다는 점에 있어서 예측 불가능하다. /어도/와 /아도/는 /ㄱ/, /ㄷ/, /ㄴ/ 이외의 모음이나 자음 뒤에서 선택되고, /도/는 모음 /ㄱ/, /ㄷ/, /ㄴ/ 직후에서 선택된다. 표에서는 이 세 변이형 가운데 하나가 요구되는 자리는 따로 표시를 하지 않았다.

첫 행은 규칙적으로 굴절하는 동작 동사들의 패러다임으로서, 어간의 끝 소리가 자음이건 모음이건 관계 없이 많은 동사들이 이 굴절 부류에 속한다. /으면/과 /면/ 등의 변이형은, /는군/과 /군/의 경우와는 달리, 그 선택이 어간의 음운적 조건에만 민감하기 때문이다. 굴절 부류 I에 속하는 동사들로는 서 'stand', 타 'obtain (ration)', 입 'wear', 접 'fold', 겪 'undergo' 등이 있는데, 구성원의 수가 가장 많은 부류다. 두 번째 행은 규칙적으로 굴절하는 상태 동사들의 패러다임이다. 이 부류 (굴절 부류 II)에 속하는 동사로는 별나 'be particular', 약 'be sly', 젖 'be wet', 짙 'be dark', 깊 'be deep' 등이 있고, 상태 동사의 굴절 부류 중에서 가장 큰 부류다.

그 다음 다섯 행의 패러다임은 모두 구성원의 수가 극히 적은 굴절 부류들의 패러다임이다. 부류 III과 IV, 그리고 V는 각각 한 개씩의 동사만을 그 구성원으로 갖는다. 굴절 부류 VI에 속하는 동사의 수는 방언에 따라 다르나 소수다.³⁾ 부류 VII에는 동사 오 'come'와 이 동사에 전속 형태소 (cranberry morph)가 접두사로 붙은 다가오 'approach', 나오 'come out' 등 극소수다.

위 표에서 부르 'beckon'와 빠르 'be fast'로 대표된 굴절 부류 VIII과 IX는 주요 어간이 /르/로 끝난 모든 동사와 (예를 들어 모르 'not know' 구르 'roll', 기르 'raise', 및 다르 'be different', 고르 'be even', 부르 '(of stomach) be full' 등) 주요 어간이 /하/나 모음 /ㅡ/, /ㅣ/, /ㅏ/ 또는 /ㅓ/ 로 끝난 동사들을 그 구성원으로 한다. 우리 말에서 가장 자주 나타나는 낱말 중의 하나인 하 'do'는 /하/와 /해/의 변이형을 갖는데, 부차어간인 /해/가 쓰이는 곳은 부르 'beckon'의 부차어간 /블르/이 쓰이는 곳과 꼭 같기 때문이다.

주요 어간의 마지막 소리가 /ㄹ/인 모든 동사는 굴절 부류 X이나 XI에 속한다. 부류 I과 II, 그리고 VIII과 IX의 대립과 같이, 부류 X과 XI의 대립은 /는군/열과 /는데/열에서 나타난다. 어간의 음운적 특징로만 굴절 부류를 예측할 수 있는 경우는 없지만, 마지막 위치에 /ㄹ/이 있느냐 없느냐로 "굴절 부류 X 또는 XI" 만큼 정확하게 예측할 수 있는 것은 우리 말 음운 형태부문이 갖는 우연한 속성이라고 본다.

굴절 부류 XII와 XIII에 속하는 동사들은 모두 어간 중 하나가 두 개의 자음 연쇄체로 끝나는 것들이다. 얹 'put (on a surface higher than the floor)', 끊 'cut', 밟 'tread' 등의 동작 동사가 부류 XII에, 굵 'be thick', 짧 'be short' 등의 상태 동사가 부류 XIII에 속한다.⁴⁾

마지막 굴절 부류쌍인 XIV와 XV는 어간을 세 개씩 갖는 동사들을 그 구성원으로 한다. 전통 문법에서 "ㄷ-변칙", "ㅅ-변칙", 그리고 "ㅂ-변칙"으로 분류되는 동사들이 다 여기에 속하고, 그 밖에도 주요 어간의 마지막 소리가 /ㅎ/인 동사가 모두 이 무리에 속한다.

3. 강 승식 (1993: 63)에 따르면 '거라' 불규칙에 속하는 용언이 모두 60개, '너라' 불규칙에 속하는 용언이 41개라 한다.

4. 동사 끊의 굴절 끝은 /끊고/, /끈으면/, /끈으시/, /끈어도/, /끊는군/, /끊는데/이다. 발음 되지 않는 /ㅎ/은 자동 음운과정으로 삭제된다고 가정한다. 현행 한글 표기법은 /ㅎ/삭제의 환경이 더 넓은 것으로 전제하고 끊, 끊 등의 동사를 모두 규칙 동사로 취급한다.

부류 XIV에 속하는 동사 넣 'insert'는 주요 어간 /넣/ 이외에, /너으/와 /너어/의 부차 어간을 갖는다. 이와 유사하게, 부류 XV에 속하는 상태 동사 그럴 'be like that'는 주요 어간으로 /그렇/을, 부차 어간으로 /그러/와 /그래/를 갖는다.

IV. 결론 및 형태소 분석에의 응용

앞 절에서 얻은 열 다섯 개의 우리 말 동사의 굴절 부류를 이용하면 I절에서 든 형태적 성과 분석의 난제들 가운데에서 첫 세가지를 해결할 수 있다. 접사의 끝 선택이 원격한 어간에 민감한 현상 (/잡는군/, /잡으시는군/, */잡으시군/; /바쁘군/, */바쁘시는군/, /바쁘시군/)에 관한 논의로는 [6]이 있다. 참고로, 이 현상은 용언의 어미를 선어말 어미와 어말 어미로 구분하고 어간과 어미 사이의 결합 가능성을 음운적 정보만으로부터 예측하려고 하는 [1], [2] 등에서는 아예 다루지 않을 만큼 어려운 현상이다.

이빨 빠진 패러다임 (defective paradigm)을 노정하는 불구동사들은 결여하는 굴절 끝의 분포에 따라 새로운 굴절부류들에 할당 될 수 있다. 불구 동사 하나하나가 독특한 방식으로 굴절 끝을 결여함으로써 독자적으로 굴절 부류를 형성하기 쉬운 듯하다. I절에서 언급한 난제들 중 마지막 항이었던 누적적 실현 문제는 실현 규칙이 “형태·통사적 속성들의 집합”을 언급하게 함으로써 해결된다.

마지막으로 이 논문에서 채택한 낱말·패러다임 형태 이론이 요즈음 널리 통용되고 있는 한국어 형태소 분석 모델들과 어떻게 다른지에 대해 알아 보자.

첫째, 이 모델은 생성 모델이다. 어휘소와 거기에 추가된 형태·통사적 속성들로부터 음운연쇄체를 얻어 낸다. 따라서 우리 말을 목표 언어로 하는 자동 번역 시스템이나 기타 한국어 문장을 생성하는 전문가 시스템이 포함 해야 할 부문이다.

둘째, 이 모델은 표기법 또는 정서법에 바탕을 두지 않고, 의미로부터 실제 발음을 어떻게 가장 경제적으로, 또 완벽하게 얻어 내느냐에 치중한다. (특정한 코딩스킴을 전제하지도 물론 않는다.) 표기법에 맞는 표기는 이 모델이 생성하는 음운 표지 (phonemic representations)에서 도출 되어야 하는 반면, 말 소리 생성을 위해서는 음성 구현 규칙들이 음운 표지에 적용 되어야 한다.

셋째, 이 모델은 한국어 뿐 아니라 굴절을 갖는 모든 언어의 기술에 이용될 수 있는 형태 부문에 관한 일반 이론이다. 희랍어와 라틴어의 문법 전통에서는 굴절 부류의 설정이 이미 수 세기 전에 설정 되어 형태 부문 기술에 쓰여 왔으나 국어 문법의 전통에서는 굴절 부류라는 개념이 없었으므로 이 논문에서 논의한 것이다.

IA-모델(Item-and-Arrangement model) 및 IP-모델(Item-and-Process model)과 대조 되어 WP-모델로 불리는 이 이론이 형태소 분석에 어떻게 적용 될지에 관해서는 발표된 논문이 없으나, 접사 추출(affix-stripping)에 의해 과다 수용된 오분석들을 걸러 주는 역할을 할 수

있음에는 틀림 없다. 이 모델을 이용한 형태소 분석의 알고리즘은 현재 개발 중에 있으므로 그 소개는 다음 기회로 미룬다.

참고 문헌

- [1] 강 승식, 음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석, 서울대학교 대학원 컴퓨터 공학과 박사 학위 논문, 1993.
- [2] 이 은철, 이 중혁, “계층적 기호 접속정보를 이용한 한국어 형태소 분석기의 구현”, 제 4회 한글 및 한국어 정보처리 학술대회 논문집, 1992.
- [3] Anderson, Stephen R., “Where’s morphology?”, *Linguistic Inquiry* 13, 571-612, 1982.
- [4] Matthews, Peter H., *Inflectional Morphology*, Cambridge University Press, Cambridge, 1972.
- [5] Matthews, Peter H., *Morphology: an Introduction to the Theory of Word-Structure*, Cambridge University Press, Cambridge, 1974 [Revised edition 1992]
- [6] No, Yongkyoon, “Refining a Word-and-paradigm morphology for Korean verbs”, Paper presented at the 1992 Autumn Meeting of the Linguistic Society of Great Britain, University of Surrey, 1992.
- [7] Zwicky, Arnold M., “How to describe inflection”, *Berkeley Linguistic Society Proceedings* 11, 372-386, 1985.