

한국어 체언말 저해음 변이를 위한 복수기저형 확률문법*

김선희**
(중앙대학교)

김하나***
(중앙대학교)

Kim, Sun-Hoi and Hana Kim. 2016. A probabilistic grammar with multiple underlying representations for Korean noun-final obstruent variation. *Studies in Phonetics, Phonology and Morphology* 22.2. 207-244. This paper aims to propose a probabilistic grammar with multiple underlying representations (UR's) for Korean noun-final obstruent variation and to analyze the influence of UR choice on this variation. The grammar is based on Pater et al.'s (2012) multiple UR Maximum Entropy model, where UR constraints are proposed to require a particular UR corresponding to the meaning/concept of a given morpheme or word. It consists of weighted constraints learned from the training data which are a result of the picture-presented production experiment conducted for this variation. This grammar produces the probabilistic distribution of variation for Korean obstruent-final nouns. The probabilistic distribution tells us that this variation consists of three different types of variation. One is a (traditional or etymological) UR-preferred type of variation (c^h -final and p^h -final nouns), another is a substitution-preferred one (k^h -final nouns) and the other is the variation which a phonological process strongly influences (t^h -final nouns). This phenomenon shows that the variation is difficult to explain without considering the UR factor. Further, in this paper, it is shown that the UR choice is determined by the interaction between UR constraints, faithfulness constraints and markedness constraints. (Chung-Ang University)

Keywords: Korean, noun-final obstruent variation, UR constraints, probabilistic grammar, multiple UR Maximum Entropy model, UR choice

* 이 논문을 심사한 익명의 세 분 심사자에게 감사 드린다. 심사자들은 논문 초고에 나타난 오류와 실수를 적절하게 지적하였고, 제외, 첨가, 보완해야 할 점들에 대해 타당한 의견들을 제시하였다. 초고에 비해 논문의 완성도가 높아졌다면, 이는 전적으로 이러한 지적과 의견들 덕분이다. 여전히 남아 있는 실수와 오류, 미진한 부분에 대한 책임은 모두 저자들의 몫임을 밝힌다.

** 제1저자, 교신저자

*** 제2저자

1. 서론

일반적으로, ‘낫’의 말음은 모음 앞에서 [cʰ]로 실현된다. 한편, ‘밭’의 말음은 -ɪ를 제외한 모음 앞에서는 [tʰ]로 실현되고 -ɪ 앞에서는 [cʰ]로 실현된다. 그리고 이 두 체언의 말음은 독립된 환경에서는 [t]로 실현된다. 전통적으로, 이 체언 말음들의 기저형은 각각 /cʰ/와 /tʰ/이고, -ɪ 앞에서 /tʰ/ → [cʰ] 구개음화, 독립된 환경에서는 /cʰ, tʰ/ → [t] 중화가 일어난다고 보았다. 그러나 독립된 환경에서는 발생하지 않으나 모음 앞 환경에서는 빈번하게 발생하는 (1)과 같은 음운변이 때문에 이 일반화는 한계를 지닌다: ‘잎’과 ‘부엌’에서도 (1b)와 같은 모음 앞 음운변이 현상이 빈번하게 발생한다.

(1) 모음 앞 체언말 저해음 변이

a. ‘낫’과 ‘밭’

	낫 ‘face’	밭 ‘field’	
-ε	nac ^h ε~nase~nat ^h ε	pac ^h ε~pase~pat ^h ε	Locative
-ɪ	nac ^h ɪ~nasi~(nat ^h ɪ)	pac ^h ɪ~pasi~(pat ^h ɪ)	Nominative
-il	nac ^h il~nasil~nat ^h il	pac ^h il~pasil~pat ^h il	Accusative

b. ‘잎’과 ‘부엌’

	잎 ‘leaf’	부엌 ‘kitchen’	
-ε	ɪp ^h ε~ɪpε	puək ^h ε~puəkε	Locative
-ɪ	ɪp ^h ɪ~ɪpɪ	puək ^h ɪ~puəkɪ	Nominative
-il	ɪp ^h il~ɪpil	puək ^h il~puəkil	Accusative

최혜원(2004)과 강은지 외(2004)를 비롯한 선행연구들에 따르면, 이 음운변이는 말음을 기준으로 구분되는 체언의 유형에 따라 변이형들의 분포가 다르다. 예를 들면, 전통적으로 말음의 기저형이 /cʰ/라고 여겨진 체언들의 경우에는 (이 연구에서는 이 체언들을 cʰ-말음 체언이라고 부를 것이다), [...cʰε] 변이형의 빈도가 [...tʰε] 변이형의 빈도보다 더 높다: ‘낫-에’가 [nac^hε]로 실현된 경우의 빈도가 [nat^hε]로 실현된 경우의 빈도보다 더 높다. 그러나 전통적으로 말음의 기저형이 /tʰ/라고 여겨진 체언들의 경우에는 (이 체언들을 tʰ-말음 체언이라고 부를 것이다), 반대로 [...tʰε] 변이형의 빈도가 [...cʰε] 변이형의 빈도보다 더 높다: ‘밭-에’가 [pat^hε]로 실현된 경우의 빈도가 [pac^hε]로 실현된 경우의 빈도보다 더 높다. 두 유형의 체언들 모두 말음이 [s]로 실현된 경우([nase], [pase])의 빈도가 무시할 수 없을 정도라는 점도 주목할 만한 사항이다.

체언의 유형에 따른 변이형들의 분포 차이는 전통적으로 발음의 기저형이 /p^h/라고 여겨진 체언들과 /k^h/라고 여겨진 체언들 사이에서도 관찰된다: 이 연구에서는 이 체언들을 각각 p^h-말음 체언, k^h-말음 체언이라고 부를 것이다. ‘잎-에’가 [ip^hɛ]로 실현된 경우의 빈도가 [ipɛ]로 실현된 경우의 빈도보다 더 높지만, ‘부엌-에’가 [puəkɛ]로 실현된 경우의 빈도가 [puək^hɛ]로 실현된 경우의 빈도보다 더 높다: 모음 사이에서 일어나는 연폐쇄음(lenis stop)의 유성음화는 이 연구의 관심 사항이 아니므로 이 연구에서는 이를 고려하지 않고 논의를 진행할 것이다. 한편, 후행모음에 따른 변이형들의 분포 차이도 관찰된다. ^h-말음 체언의 경우 -ɛ 모음 앞 [t^h]-변이형의 빈도는 -i 모음 앞 [t^h]-변이형의 빈도보다 훨씬 높은 반면, [c^h]-변이형의 빈도는 반대로 -ɛ 모음 앞에서보다 -i 모음 앞에서 훨씬 높다(최혜원 2004, 강은지의 2004).

이 연구는 위에서 기술한 한국어 체언말 저해음 변이의 패턴과 분포편향성, 변이형의 확률분포를 확률적 접근을 통해 분석함으로써, 이 변이와 관련된 한국어 화자의 문법을 제시하는 것을 목적으로 한다. 여러 선행연구들은 화자들 사이의 기저형 설정의 차이 즉, 동일 단어에 대한 복수기저형의 존재가 음운변이를 초래하는 중요한 요인들 가운데 하나라고 추론한다(Guy 2003: 379). 그러나 추론에 그칠 뿐, 동일 단어에 대한 복수기저형들의 확률분포와 해당 단어의 변이형들의 확률분포 사이의 관계를 분석한 연구들은 매우 드물다. 그 이유는 동일 단어의 복수기저형들의 분포를 확률적으로 도출하는데 어려움이 있기 때문이다. 한국어 체언말 저해음 변이에 대한 선행연구들에서도 동일 단어에 대한 복수기저형들의 확률분포를 도출하고자 시도한 분석은 발견되지 않는다.

이 연구는 한국어 화자들이 동일 단어에 대해 서로 다른 기저형을 가질 수 있으며, 이 기저형 선택의 차이가 체언말 저해음 변이를 일으키는 요인들 가운데 하나라는 복수기저형 관점에 입각한다. 그리고 관찰된 데이터에 나타난 변이형들의 확률분포로부터 복수기저형들의 확률분포를 도출하는 Pater et al. (2012)의 복수기저형 Maximum Entropy 문법(이하, MaxEnt문법)을 채택한다. 따라서 이 연구의 분석은 가중치 제약(weighted constraints)을 가진 MaxEnt문법(Johnson 2002, Goldwater and Johnson 2003, Wilson 2006, Jäger 2007)에 바탕을 둔다.

복수기저형 관점에 따르면, /c^h/뿐 아니라 /s/와 /t^h/도 c^h-말음 체언의 발음 기저형이 될 수 있다. ^h-말음 체언의 발음 기저형 역시 /c^h/, /s/, /t^h/ 모두 가능하다. 다시 말해서, 화자들은 c^h-말음 체언과 ^h-말음 체언의 발음 기저형으로 이 가운데 하나 또는 복수의 기저형을 선택한다. 그렇다고 해서 발음

을 기준으로 한 체언들의 유형 구분이 무의미한 것은 결코 아니다. 이 연구는 관찰된 데이터에 나타난 변이형들의 확률분포에 근거하여 MaxEnt 방법이 복수기저형들의 확률분포를 도출할 때 각 체언 유형별로 다른 확률분포가 도출된다는 것을 보임으로써, 말음을 기준으로 한 체언들의 유형 구분이 필수불가결함을 보일 것이다. 이 연구에서 체언들의 유형 구분은 특정 기저형이 체언의 의미/개념에 연결될 것을 요구하는 기저형 제약들(이하, UR제약)의 가중치를 정하는데 결정적인 역할을 한다: 이 제약은 Pater et al. (2012)이 불어 모음탈락 변이의 분석에서 제안한 UR제약과 동일한 특성을 갖는 제약이다.

복수기저형 관점에 따르면, 의미/개념 x 가 변이형 y_1, y_2, \dots, y_n 으로 실현될 때, 이들을 연결하는 것은 복수기저형 z_1, z_2, \dots, z_n 이다. 충실성제약과 유표성제약이 음운과정 $\{z_1, z_2, \dots, z_n\} \rightarrow \{y_1, y_2, \dots, y_n\}$ 을 평가하는 것처럼, 기저형 선택 즉, $x \rightarrow z_1, x \rightarrow z_2, \dots, x \rightarrow z_n$ 을 평가하는 UR제약이 존재한다. 예를 들면, c^h -말음 체언의 의미/개념 x_1 이 모음 앞에서 변이형 [...c^h], [...s], [...t^h]로 실현될 때, 이들을 연결하는 것은 복수기저형 /...c^h/, /...s/, /...t^h/이다. 또한, t^h -말음 체언의 의미/개념 x_2 가 모음 앞에서 변이형 [...c^h], [...s], [...t^h]로 실현될 때, 이들을 연결하는 것 역시 복수기저형 /...c^h/, /...s/, /...t^h/이다. c^h -말음 체언의 의미/개념 x_1 의 기저형으로 /...c^h/가 선택될 것을 요구하는 제약을 UR₁이라고 하고, t^h -말음 체언의 의미/개념 x_2 의 기저형으로 /...c^h/가 선택될 것을 요구하는 제약을 UR₂이라고 하면, UR₁과 UR₂의 가중치는 데이터에 나타난 c^h -말음 체언의 변이형들의 확률분포와 t^h -말음 체언의 변이형들의 확률분포에 따라 달리 결정된다. 따라서 복수기저형을 가정하고 있을지라도 말음을 기준으로 한 체언 유형의 구분은 이 연구에서 UR제약들의 가중치를 정하는데 결정적인 역할을 한다.

한편, 기저형 /...c^h/는 변이형 [...c^h]뿐 아니라 [...s] 또는 [...t^h]로도 실현될 수 있다. 이 때 각 변이형의 실현확률은 데이터에 나타난 c^h -말음 체언의 변이형들의 확률분포와 t^h -말음 체언의 변이형들의 확률분포에 따라 가중치가 결정된 충실성제약들과 유표성제약들의 상호작용에 의해 결정된다. 결국, 이 연구에서는 의미/개념 x 에 대한 기저형-변이형쌍(UR-variant pair) 즉, $x-z_1-y_1, x-z_1-y_2, \dots, x-z_1-y_n$ 과 $x-z_2-y_1, x-z_2-y_2, \dots, x-z_2-y_n, \dots, x-z_n-y_1, x-z_n-y_2, \dots, x-z_n-y_n$ 각각의 확률을 가중치를 가진 UR제약, 충실성제약, 유표성제약들의 상호작용에 의해 도출한다. 즉, 각 변이형에 대한 복수기저형들의 확률분포를 도출함으로써 기저형 선택이 변이형 분포패턴에 끼치는 영향 정도를 계량화한다.

이 글의 구성은 다음과 같다. 2절에서는 이 연구가 수행한 발화실험의

내용과 방법, 결과를 제시한다. 3절에서는 MaxEnt문법에 복수기저형 체계를 결합시킨 문법체계에 대해 토론했다. 4절에서는 이 문법에 입각한 한국어 체언말 저해음 변이에 대한 분석을 제시하고 분석 결과에 대해 토론했다. 5절에서는 이 연구의 분석 결과와 다른 연구들의 분석 결과를 비교한다. 결론은 6절에 제시된다.

2. 발화실험의 내용과 결과

2.1 방법과 내용

방언/연령 차이에 따른 변이패턴의 차이를 배제하고자 참여대상을 표준한국어 음운특징을 보이는 서울/인천/경기 지역에서 출생·성장한 20대로 제한하고, 한 대학교의 온·오프라인 게시판을 통해 이 지역에서 출생·성장한 20대 참여자 44명을 모집하였다: 남성 22명, 여성 22명, 서울 47.7%, 인천 27.3%, 경기 25%. 연령 만 20세~27세, 평균 연령 만 22.6세. 모두 자발적 참여의사를 나타냈으며, 실험후 소정의 물질적 보상이 제공되었다. 7분~10분 가량 개별적으로 이루어진 일상적 대화를 통해 이들 모두 한국어 발화 수행에 이상이 없음을 확인하였다. 이 대화내용은 녹취되지 않았다.

c^h -말음 체언, t^h -말음 체언, p^h -말음 체언, k^h -말음 체언들 중 일부를 실험대상어로 선택하였다. 『우리말 역순 사전』(유재원 1985)에 제시된 해당 체언 469개 가운데 (c^h -말음 체언 172개, t^h -말음 체언 184개, p^h -말음 체언 96개, k^h -말음 체언 17개), 30개의 체언을 선택하였다: c^h -말음 체언 10개, t^h -말음 체언 9개, p^h -말음 체언 9개, k^h -말음 체언 2개¹.

한글철자의 영향으로 실험결과가 왜곡될 가능성을 배제하기 위해 이미지를 이용한 발화실험을 수행하였다. 따라서 시각화의 용이성 여부가 대상

¹ 『우리말 역순 사전』에 등재된 k^h -말음 체언들은 ‘부엌’이 포함된 합성어 2개와 ‘부엌’, ‘뽕(부엌의 줄임말)’과 ‘넋’으로 끝나는 단어 12개, 그리고 ‘키읔’이었다. ‘키읔’은 철자의 영향을 고려하여 제외하였고 ‘부엌’과 ‘넋’으로 끝나는 체언 가운데 ‘서넋’을 실험대상어로 선택하였다. 이 때문에 k^h -말음 체언이 2개밖에 선택되지 않는 편중성이 발생하였다. 선정된 실험대상어들은 다음과 같다. c^h -말음 체언: 꽃, 낫, 닛, 텃, 돛, 빗, 살갓, 솟, 옷, 옷, t^h -말음 체언: 꽃밭, 끝, 뒤꼍, 물, 밭, 바깥, 술, 팔, 햇볕, p^h -말음 체언: 길썬, 깎았, 늙, 대나무숲, 무릎, 벗졌, 앞, 옆, 형겼, k^h -말음 체언: 부엌, 서넋.

어 선택에 영향을 끼쳤다². 실험은 2015년 11월 9일~12월 25일 사이에 음성녹음실에서 참여자 개인별로 진행되었고 실험 진행방식은 다음과 같다.

목표어를 유도하는 이미지와 함께 목표어 자리가 비어있는 문장이 컴퓨터 화면을 통해 참여자에게 제시되었다. 참여자에게 제시된 화면은 <그림 1>과 같다.

항해 전에 올려야 하는 것은 []이다.



오솔길을 []에서 발견했다.

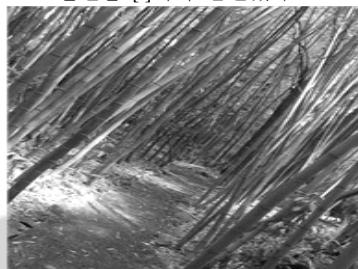


그림 1. 실험 문장과 그림의 예

참여자는 제시된 화면에서 이미지를 보고 이미지가 의미하는 체언을 빈 칸에 삽입하여 발화하였다. 후행형태소는 -ε (에), -εsə (에서), -εnin (에는), -εge (에게), -ida (이다), -il (을)로, 후행모음 환경을 /ε, ɪ, i/로 제한하였다. 각 음운환경에서 체언이 세 번씩 발화되게 실험을 설계하였다. 따라서 각 참여자가 발화한 문장 수는 총 270개였다: 체언 30개, 후행모음 3개, 발화횟수 3회. 그림이 유도하지 않은 체언을 발화한 경우와 그림에 적절한 체언을 알려 줄 것을 요청한 경우가 때때로 발생하였는데(실험 참여자 44명이 발화한 횟수 총 11,880회 가운데 538회(4.53%)), 이 경우 진행자는 그림에서 사물을 손가락으로 짚어 주었고, 그래도 적절한 체언을 찾지 못하면 해당 체언을 단독형으로 발화하여 알려 주었다. 그림은 모두 저작권을 침해하지 않는 범위에서 온라인을 통해 구했으며, 모든 발화는 Philips사의 VTR5100모델 녹음기로 녹음되었다.

² 각 체언에 해당하는 한자어를 가짜문항으로 하여 한자어-순우리말 연결에 대한 사전 테스트를 수행하여 참여자가 이 실험의 목적을 알아채지 않도록 하는 데에 주의를 기울였다.

2.2 실험결과

발화실험에서 나타난 변이형들의 빈도는 <그림 2>와 같다.

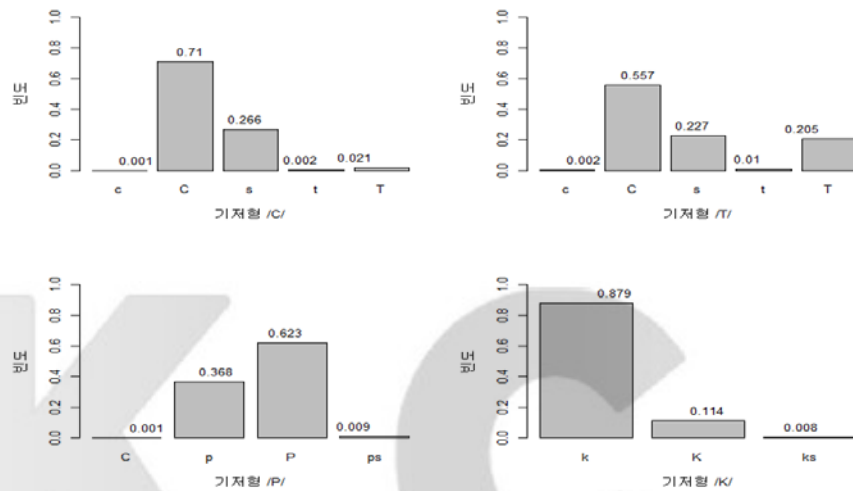


그림 2. 변이형의 빈도

<그림 2>에서 대문자 C, T, P, K는 각각 c^h , t^h , p^h , k^h 를 나타낸다. 따라서 기저형 /C/, 기저형 /T/, 기저형 /P/, 기저형 /K/는 각각 c^h -말음 체언, t^h -말음 체언, p^h -말음 체언, k^h -말음 체언을 의미한다. 이 실험의 결과는 체언 유형에 따라 변이형들의 분포패턴이 달라지는 것을 보여 준다. c^h -말음 체언과 p^h -말음 체언, k^h -말음 체언의 경우 말음이 전통적 기저형이 아닌 다른 저해음으로 발화될 때에는 각각 [s, p, k]로 실현되는 것이 일반적이다: 이를 대체형이라 부를 것이다. c^h -말음 체언과 p^h -말음 체언의 경우에는 전통적 기저형의 실현 빈도가 더 높은 반면, k^h -말음 체언에서는 대체형의 실현 빈도가 더 높다는 점이 특징적이다.

표 1. 후행모음에 따른 변이형의 분포³

(%)	<i>c^h</i> -말음 체언			<i>t^h</i> -말음 체언			<i>p^h</i> -말음 체언			<i>k^h</i> -말음 체언			
	-ε	-ɪ	-i	-ε	-ɪ	-i							
[c ^h]	67.7	70.8	74.9	29.6	71.5	65.9							
[s]	26.6	28.4	24.9	18.8	25.8	23.6							
[t ^h]	5.5	0.2	0.6	48.4	2.4	10.5							
							[p ^h]	65.3	59.0	62.5			
							[p]	34.5	39.7	36.1			
										[k ^h]	14.8	9.1	10.2
										[k]	85.2	89.8	88.6

<표 1>에 따르면, *c^h*-말음 체언, *p^h*-말음 체언, *k^h*-말음 체언들의 변이형 분포패턴은 후행모음과 관계없이 일정하다. *c^h*-말음 체언에서는 후행모음과 관계없이 [c^h] > [s] > [t^h] 순이며, [c^h]의 실현 빈도가 가장 높은 반면, [t^h]의 실현 빈도는 매우 낮다. 후행모음과 관계없이 *p^h*-말음 체언에서는 전통적 기저형의 실현 빈도, *k^h*-말음 체언에서는 대체형의 실현 빈도가 더 높다. 그러나 *ɸ^h*-말음 체언의 변이형 분포패턴은 후행모음에 따라 다르다. 주목할 점은 구개음화 적용환경인 -ɪ 모음뿐 아니라 -i 모음 앞에서도 [c^h] > [t^h] 순인데 반해, -ε 모음 앞에서는 [t^h]로 실현되는 빈도가 더 높다는 것이다. 이것은 체언말 저해음의 변이패턴을 분석하는데 있어서 *c^h*-말음 체언과 *ɸ^h*-말음 체언이 구분되어 취급되어야 함을 보여 준다. 다음 절에서는 이 연구의 문법틀인 복수기저형 MaxEnt문법의 이론적 토대에 대해 토론한다.

3. 복수기저형 MaxEnt문법

3.1 MaxEnt문법

MaxEnt문법은 정보이론(Information Theory) (Shannon 1948)에서 시스템 정보량을 나타내는 엔트로피를 확률적 최적성이론(Probabilistic Optimality Theory)에 적용한 문법이다(홍성훈 2014: 61). 이 문법은 로그-선형확률모델(log-linear probabilistic model)을 기반으로 한다. 이 문법은 학습데이터(training data)의 변이형 분포패턴에 근접한 확률분포들 가운데 학습데이터로 알 수 있는 정보에 대해서는 그 최대치를 반영하고 학습데이터에 나타나지 않은 불필

³ 극히 적은 빈도를 나타낸 변이형들은 <표 1>에 표시하지 않았다.

요한 정보는 최대한도로 배제한 확률분포를 최종 확률분포로 도출한다 (Goldwater and Johnson 2003). 따라서 이 문법이 도출한 확률분포는 학습데이터에 나타난 변이형들의 빈도분포와 매우 유사하지만 정확하게 일치하지는 않는다. 이 문법은 가중치를 갖는 제약들로 구성되고, 이 가중치는 학습데이터 빈도분포를 학습한 결과이다. 확률분포는 식(2)에 의해 도출된다(Goldwater and Johnson 2003: 112).

$$(2) \text{ a. } Z(x) = \sum_{y \in Y(x)} \exp\left(\sum_{i=1}^m w_i f_i(y, x)\right) \text{ 일 때,}$$

$$\text{ b. } Pr(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{i=1}^m w_i f_i(y, x)\right)$$

$Pr(y | x)$ 는 x 에 대한 y 의 조건부 확률(conditional probability)로, 입력형 x 가 출력형(변이형) y 로 실현될 확률이다. $f_i(y, x)$ 는 출력형-입력형쌍 함수이다. 이것은 입력형이 x 일 때 출력형 y 가 제약들을 위반한 횟수이다. w_i 는 제약의 가중치이고, m 은 제약의 개수이다. 따라서 (2a)의 $\sum_{i=1}^m w_i f_i(y, x)$ 는 입력형이 x 일 때 출력형 y 가 각 제약을 위반한 횟수와 해당 제약의 가중치를 곱한 값을 모두 더한 값이다. 이를 조화치(harmony value, 이하, H-value)라고 한다. 그리고 $\exp(\sum_{i=1}^m w_i f_i(y, x))$ 는 H-value를 지수(exponent)로 전환한 값이다: 편의상, 이 값을 (y, x) 의 e^H 라고 할 것이다. $Z(x)$ 는 입력형 x 로부터 도출될 가능성이 있는 출력형 집합을 $Y(x)$ 라 할 때, $Y(x)$ 를 구성하는 각 출력형들의 H-value를 지수로 전환한 값을 모두 더한 값이다. 따라서 조건부 확률 $Pr(y | x)$ 는 (y, x) 의 e^H 를 x 로부터 도출가능한 모든 출력형 y 들의 e^H 를 더한 값으로 나눠 산출한 값이다.

지금부터는 $Pr(y | x)$ 를 구하는데 결정적 역할을 하는 제약의 가중치 w_i 를 어떻게 산출하는지에 대해서 알아본다. w_i 는 학습데이터의 조건우도(conditional likelihood)를 최대화하는 식(3)을 가지고 도출한다(Goldwater and Johnson 2003: 113).

$$(3) \text{ a. } PL_{\bar{w}}(\bar{y} | \bar{x}) = \prod_{j=1}^n Pr_{\bar{w}}(Y = y_j | x(Y) = x_j)$$

$$\text{ b. } \log PL_{\bar{w}}(\bar{y} | \bar{x}) = \sum_{i=1}^m \frac{(w_i - \mu_i)^2}{2\sigma_i^2}$$

(3a)에서 \bar{y} 는 학습데이터의 n 개 입력형 각각에 대한 ‘승리출력형들(winning outputs)’이고, x_j 는 해당 입력형들이다. 따라서 학습데이터 조건우도는 각 입력형에 대한 출력형들의 조건부 확률을 곱한 값이다. (3b)에서는 (3a)에서 구한 조건우도를 최대화하기 위해 로그를 취하였으며 학습데이터의 과적합(overfitting)을 방지하기 위해 평균 μ_i 와 표준편차 σ_i 에 의해 결정되는 가우스분포(Gaussian distribution) (정규분포)를 사전분포(prior)로 택하였다(Goldwater and Johnson 2003: 113).

3.2 MaxEnt문법의 예측

이제, MaxEnt문법이 어떻게 확률분포를 도출하는지 예를 통해 살펴보자. ‘밭-을’이 [pas-il], [pac^h-il], [pat^h-il]로 실현되는 변이를 이용하되, 설명의 편의를 위해 다른 제약들은 고려하지 않고 유표성제약 *si, *c^hi, *t^hi에만 초점을 맞출 것이다: 여기에서는 ‘밭’의 기저형을 /pat^h/이라고 가정하였다. <표 2>의 제약 가중치(제약 밑에 적힌 수) 역시 설명의 편의를 위해 설정한 임의의 값이다: 실제 가중치는 학습데이터가 존재할 때 식(3)을 가지고 도출한다.

<표 2>에서 [pas-il], [pac^h-il], [pat^h-il]의 H-value ($\sum_{i=1}^m w_i f_i(y, x)$)는 음수로 표시된 각 출력형의 제약위반 횟수를 해당 제약의 가중치와 곱한 뒤 그 값을 모두 더한 값이다. 이 값은 각각 -5, -4, -7이다. 따라서 $\exp(\sum_{i=1}^m w_i f_i(y, x))$ 인 e^H 는 0.0067, 0.0183, 0.0009이다. 전체 출력형 집합 $Y(x)$ 의 e^H 인 $Z(x)$ 는 이들을 모두 더한 값인 0.0259이다. 각 출력형의 e^H 값을 $Z(x)$ 값 0.0259로 나누면 MaxEnt문법이 예측하는 [pas-il], [pac^h-il], [pat^h-il]의 확률이 도출된다 (<표 2>에서 p 값).

표 2. MaxEnt문법의 예

‘밭-을’ /pat ^h -il/	*si 5	*c ^h i 4	*t ^h i 7	H value	e^H	p
[pas-il]	-1			-5	0.0067	25.9
[pac ^h -il]		-1		-4	0.0183	70.6
[pat ^h -il]			-1	-7	0.0009	3.5

요약하면, MaxEnt문법에 입력되어야 할 것은 각 출력형의 제약위반 횟수를 포함한 제약위반 사항과 학습데이터의 변이형 빈도분포이다. MaxEnt문법은 학습데이터의 빈도분포에 끼친 제약들의 영향 정도를 계량화하여 제약들에게 가중치를 부여한다. 이 가중치 제약문법은 학습데이터에 포함되

어 있지 않은 데이터 즉, 학습데이터를 표본으로 한 모집단의 변이형들의 확률분포를 도출한다.

3.3 복수기저형 MaxEnt문법

<표 2>의 예는 사실상 ‘밭’의 기저형을 /pat^h/이라고 본 단일기저형 체계를 가정한 것이었다. 단일기저형 체계 대신 아래 <그림 3>과 같은 복수기저형 체계를 채택할 때에는 ?에 해당하는 <입력의미/개념→기저형>과 관련된 문법적 작용이 형식화되어야 한다.

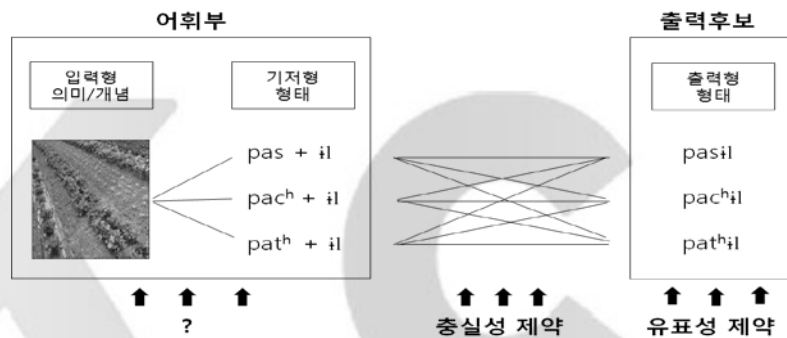


그림 3. 복수기저형 체계

1절에서 언급했듯이, 이 연구는 UR제약을 이에 해당하는 문법적 작용으로 본 Pater et al. (2012)의 제안을 채택하였다. 이 제안을 한국어 체언말 저해음 변이에 적용하면, 모음 앞에서 나타나는 변이형들 가운데 하나를 기저형으로 선택할 것을 요구하는 UR제약들이 존재하는 것으로 가정할 수 있다. ‘밭’의 의미/개념 ‘FIELD’에 대해, ‘FIELD’→/pas/, ‘FIELD’→/pac^h/, ‘FIELD’→/pat^h/과 같은 UR제약이 존재한다고 보면, 기저형 /pas/는 UR제약 ‘FIELD’→/pas/를 만족시키지만 또 다른 UR제약들 ‘FIELD’→/pac^h/와 ‘FIELD’→/pat^h/를 위반한다. 이 UR제약들은 다른 기저형 /pac^h/와 /pat^h/를 마찬가지로 방식으로 평가할 것이다.

이 체계를 MaxEnt문법에 적용하면 입력형 x 가 주어졌을 때 출력형 y 가 나올 확률을 구하는 식(2)는 x, y 이외에도 기저형 z 를 추가하는 다음 식으로 확장되어야 한다(Pater et al. 2012: 65).

$$(4) \text{ a. } Z(x) = \sum_{y \in Y(x)} \exp\left(\sum_{i=1}^m w_i f_i(y, z, x)\right) \text{ 일 때,}$$

$$\text{b. } Pr(y | x) = \frac{1}{Z(x)} \exp\left(\sum_{i=1}^m w_i f_i(y, z, x)\right)$$

입력형과 기저형을 따로 구분하지 않는 단일기저형 체계에서 $Pr(y | x)$ 는 입력형(=기저형) x 에 대한 출력형 y 의 실현확률이었다. 그러나 입력형(의미/개념)과 기저형이 따로 구분되는 복수기저형 체계에서 $Pr(y | x)$ 는 입력형(의미/개념) x 에 대한 복수기저형 z_1, z_2, \dots, z_n 에 연결된 변이형 y_1, y_2, \dots, y_n 각각의 실현확률이다. 또한, 복수기저형 $z_1, z_2, z_3 \dots z_n$ 이 존재하므로 복수기저형에 작용하는 UR제약과 충실성제약의 위반 횟수가 고려 대상에 추가되어야 한다. 이 체계에서는 충실성제약과 유표성제약뿐 아니라 UR제약들의 가중치도 산출되어야 하는데 이것은 식(3)에 의해 도출된다.

3.4 복수기저형 MaxEnt문법의 예측

이 절에서는 3.2절에서 다루었던 예가 복수기저형 체계에서는 어떻게 다르게 분석되는지 살펴본다. 앞서 설정했던 유표성제약들의 가상 가중치와 UR제약, 충실성제약들의 가상 가중치가 (5)와 같다고 가정할 것이다.

(5) a. UR제약

‘FIELD’→/pas/: 2, ‘FIELD’→/pac^h/: 3, ‘FIELD’→/pat^h/: 1

b. 충실성제약

/pas/↔[pas]: 2, /pac^h/↔[pac^h]: 2, /pat^h/↔[pat^h]: 1

c. 유표성제약

*si: 5, *c^hi: 4, t^hi: 7

입력형 ‘FIELD’-il에 대한 기저형-변이형쌍은 (6)과 같다.

(6) 입력형 ‘FIELD’-il에 대한 기저형-출력형(변이형) 쌍

변이형 [pasil]: /pas-il/-[pasil], /pac^h-il/-[pasil], /pat^h-il/-[pasil]

변이형 [pac^hil]: /pas-il/-[pac^hil], /pac^h-il /-[pac^hil], /pat^h-il/-[pac^hil]

변이형 [pat^hil]: /pas-il/-[pat^hil], /pac^h-il /-[pat^hil], /pat^h-il/-[pat^hil]

<표 3>은 복수기저형 MaxEnt문법이 변이형 확률분포를 어떻게 도출하는

지를 보여 준다: 공간상의 제약으로 하나의 표가 셋으로 나누어져 있다.

표 3. 복수기저형 MaxEnt문법의 예

‘발-을’ ‘FIELD’-il		FIELD→/pas/	FIELD→/pac ^h /	FIELD→/pat ^h /
기저형	출력형	2	3	1
/pas-il/	[pasil]		-1	-1
/pac ^h -il/	[pasil]	-1		-1
/pat ^h -il/	[pasil]	-1	-1	
/pas-il/	[pac ^h il]		-1	-1
/pac ^h -il/	[pac ^h il]	-1		-1
/pat ^h -il/	[pac ^h il]	-1	-1	
/pas-il/	[pat ^h il]		-1	-1
/pac ^h -il/	[pat ^h il]	-1		-1
/pat ^h -il/	[pat ^h il]	-1	-1	

‘발-을’ ‘FIELD’-il		/pas/↔[pas]	/pac ^h /↔[pac ^h]	/pat ^h /↔[pat ^h]
기저형	출력형	2	2	1
/pas-il/	[pasil]			
/pac ^h -il/	[pasil]		-1	
/pat ^h -il/	[pasil]			-1
/pas-il/	[pac ^h il]	-1		
/pac ^h -il/	[pac ^h il]			
/pat ^h -il/	[pac ^h il]			-1
/pas-il/	[pat ^h il]	-1		
/pac ^h -il/	[pat ^h il]		-1	
/pat ^h -il/	[pat ^h il]			

‘발-을’ ‘FIELD’-il		*s _i	*c ^h _i	*t ^h _i	<i>H</i>	<i>e^H</i>	<i>p</i>
기저형	출력형	5	4	7			
/pas-il/	[pasil]	-1			-9	0.00012	9.88
/pac ^h -il/	[pasil]	-1			-10	0.00005	4.12
/pat ^h -il/	[pasil]	-1			-11	0.00002	1.65
/pas-il/	[pac ^h il]		-1		-10	0.00005	4.12
/pac ^h -il/	[pac ^h il]		-1		-7	0.00091	74.96
/pat ^h -il/	[pac ^h il]		-1		-10	0.00005	4.12
/pas-il/	[pat ^h il]			-1	-13	0.000002	0.16
/pac ^h -il/	[pat ^h il]			-1	-12	0.000006	0.49
/pat ^h -il/	[pat ^h il]			-1	-12	0.000006	0.69

<표 3>에서 제약 가중치는 학습데이터에 근거하지 않은 임의의 값이므로 변이의 확률분포 역시 가상적이다. H-value는 세 표의 각 행에 음수로 표시된 기저형-출력형쌍의 제약위반 횟수에 각 제약의 가중치를 곱하고 그 값을 모두 더한 값이다. 첫 행 'FIELD'-il→/pas-il/→[pasil]을 예로 들어 살펴보면, 이 연쇄의 H-value는 -9이다. 이 값은 UR제약 'FIELD'→/pac^h/(3)와 'FIELD'→/pat^h/(1), 유표성제약 *si(5)를 위반한 결과이다. 모든 충실성제약과 UR제약 'FIELD'→/pas/, 유표성제약 *c^hi, *t^hi는 위반되지 않는다. H-value -9를 지수로 환산한 값인 'FIELD'-il→/pas-il/→[pasil]의 e^H 값은 0.00012이다. 전체 기저형-출력형쌍 집합 $Y(x)$ 의 e^H 인 $Z(x)$ 는 이런 방식으로 각 기저형-출력형쌍의 e^H 를 구한 뒤 모두 더한 결과이고 그 값은 0.001214이다. 따라서 'FIELD'-il→/pas-il/→[pasil]의 확률은 이 쌍의 e^H 값 0.00012를 $Z(x)$ 값 0.001214로 나눈 결과인 9.88%이다.

<표 3>에 따르면, (5)의 임의의 제약 가중치를 가정할 때 'Field'가 -il 앞에서 [pas]으로 실현될 확률은 15.65%이고 [pac^h]으로 실현될 확률은 83.2%, [pat^h]으로 실현될 확률은 1.34%이다. 이 확률분포는 한 입력형을 위해 선택된 서로 다른 기저형들로부터 각 출력형이 실현된 확률을 더한 값이라는 것을 (7)에서 확인할 수 있다.

(7) '밭-을' 'FIELD'-il의 변이형들의 실현확률 (<표 3> 가상 예)

- a. 'FIELD'-il → [pasil]: 15.65%
 'FIELD'-il → /pas-il/ → [pasil]: 9.88%
 'FIELD'-il → /pac^h-il/ → [pasil]: 4.12%
 'FIELD'-il → /pat^h-il/ → [pasil]: 1.65%
- b. 'FIELD'-il → [pac^hil]: 83.2%
 'FIELD'-il → /pas-il/ → [pac^hil]: 4.12%
 'FIELD'-il → /pac^h-il/ → [pac^hil]: 74.96%
 'FIELD'-il → /pat^h-il/ → [pac^hil]: 4.12%
- c. 'FIELD'-il → [pat^hil]: 1.34%
 'FIELD'-il → /pas-il/ → [pat^hil]: 0.16%
 'FIELD'-il → /pac^h-il/ → [pat^hil]: 0.49%
 'FIELD'-il → /pat^h-il/ → [pat^hil]: 0.69%

(7)은 각 변이형이 어떤 기저형들로부터 어느 정도의 확률을 가지고 도출되는지를 보여 준다. 예를 들면, [pasil]로 실현된 15.65%는 /pas-il/→[pasil] 9.88%, /pac^h-il/→[pasil] 4.12%, /pat^h-il/→[pasil] 1.65%의 합이다. 따라서 [pasil]

로 실현된 경우의 63.13%가 -i 모음 앞에서 기저형 /pas/이 그대로 실현된 경우이다. 그러나 나머지 36.87%는 다른 기저형 즉, /pac^h/ (26.33%) 또는 /pat^h/ (10.54%)으로부터 도출된 경우이다.

지금까지 살펴본 예는 복수기저형 MaxEnt문법이 학습데이터로부터 UR 제약과 충실성제약, 유표성제약들의 가중치를 정확하게 학습할 수만 있다면 변이형의 확률분포뿐 아니라 각 변이형이 어떤 기저형으로부터 어느 정도의 확률을 가지고 도출되는지를 예측할 수 있음을 보여 준다. 이 문법의 장점은 학습데이터에서 겉으로는 드러나지 않는 기저형 선택의 확률을 학습데이터의 빈도분포를 가지고 도출할 수 있다는 것이다. 4절에서는 2절에서 제시된 발화실험 결과를 학습데이터로 삼아 한국어 체언말 저해음 변이에 관여하는 제약들의 가중치를 구하고, 변이형 확률분포와 각 기저형-변이형쌍의 확률인 기저형 선택의 확률을 예측하는 분석이 제시된다. 이 분석은 Pater et al. (2012)이 블로그(<http://blogs.umass.edu/hgr/hg-in-r/>)를 통해 공개한 R-code들(Staub 2011) 가운데 일부를 R (R Development Core Team 2015)에서 구현하는 방식을 통해 이루어졌다.

4. 분석과 토론

4.1 체언말 저해음 변이를 위한 확률문법과 확률분포

c^h-말음 체언과 *t^h*-말음 체언의 기저형-변이형쌍은 모두 54개이다: 체언 유형 2개 (*c^h*-말음 체언과 *t^h*-말음 체언), 기저형 3개 (/c^h-, /s-, /t^h-기저형), 변이형 3개 ([c^h]-변이형, [s]-변이형, [t^h]-변이형), 후행모음 3개 (-ε, -i, -i 모음). 이 분석에서는 이 체언들의 변이에 관여하는 제약 18개를 아래 (8)과 같이 설정하였다.

(8) 제약 (*c^h*-말음 체언과 *t^h*-말음 체언의 변이)

a. UR 제약 (6개)

c^h-말음 체언: {c^h}→/c^h/, {c^h}→/s/, {c^h}→/t^h/

t^h-말음 체언: {t^h}→/c^h/, {t^h}→/s/, {t^h}→/t^h/

예; {c^h}→/c^h/

c^h-말음 체언의 기저형은 /c^h/로 끝난다.

b. 충실성 제약 (3개)

Faith-*c^h* (/c^h↔[c^h]), Faith-*s* (/s↔[s]), Faith-*t^h* (/t^h↔[t^h])

예; Faith-*c^h*

/c^h/로 끝나는 기저형의 음성형태는 [c^h]로 끝난다.

c. 유표성 제약 (9개)

*c^hɛ, *c^hi, *c^hɪ, *sɛ, *sɪ, *t^hɛ, *t^hi, *t^hɪ

예; *c^hɛ: [c^hɛ]의 실현 금지

단어간 변이까지 고려한 확률문법을 제시하려면 UR제약은 훨씬 더 많아 야 하지만, 이 연구에서는 단어간 변이를 고려하지 않은 확률문법만을 제시하고자 한다. 따라서 목표단어들을 모두 개별적으로 다루기 보다는 c^h-말음 체언들을 하나로 묶고 t^h-말음 체언들을 다른 하나로 묶어 분석할 것이다.

c^h-말음 체언과 t^h-말음 체언들이 각각 /c^h/ 또는 /s/, /t^h/로 끝나는 기저형을 택할 수 있으므로 6개의 UR제약이 제안되었다. 3개의 충실성제약은 각 기저형의 말음 /c^h, s, t^h/가 출력형에 그대로 실현될 것을 요구하는 제약들이다. 유표성제약은 후행모음 -ɛ, -ɪ, -i와 결합할 때 출력형에 실현되는 음의 유표성을 평가하는 제약이다. 순/연구개저해음으로 끝나는 체언말 저해음의 변이형 [p, p^h, k, k^h]에 대해서도 (8)과 유사한 제약들을 설정하였다. R-code에 의해 MaxEnt문법이 학습데이터의 변이형 빈도분포와 제약위반 사항을 학습한 후 도출한 저해음 변이와 관련된 제약 가중치는 (9)와 같다⁴.

(9) 제약 가중치

a. UR제약

설정저해음:

{c^h}→/c^h/ (2.66) } {c^h}→/s/ (2.20) } {c^h}→/t^h/ (0.01)

{t^h}→/t^h/ (2.58) } {t^h}→/c^h/ (0.86) } {t^h}→/s/ (0.03)

순/연구개저해음:

{p^h}→/p^h/ (1.20) } {p^h}→/p/ (0.80)

{k^h}→/k/ (1.52) } {k^h}→/k^h/ (0)

⁴ 심사자 가운데 한 분은 s-말음 체언이 c^h-말음 체언과 t^h-말음 체언의 변이에 끼치는 영향을 고려해야 함을 지적하였다. s-말음 체언의 말음이 모음 앞에서 거의 예외없이 [s]로 실현된다고 가정한다면, 이 체언의 말음 기저형을 /s/로만 보아도 무방할 것이다. 다만, 이 체언의 영향으로 인해 제약 Faith-s의 가중치가 더 커질 가능성이 있고, Faith-s의 가중치가 커지면 다른 제약들의 가중치가 조정될 것이다. 제약들이 상호작용 하기 때문에 조정의 결과가 전체 결과에 영향을 끼치지 않는 것으로 예상되지만 이 연구에서는 이를 분석하지 않았다. 이 지적은 후속 연구의 필요성을 제기한다.

b. 충실성 제약

설정저해음: Faith- c^h (3.11) } Faith- s (2.52) } Faith- t^h (1.72)순저해음: Faith- p^h (1.12) } Faith- p (0.80)연구개저해음: Faith- k (1.52) } Faith- k^h (0.47)

c. 유표성 제약

설정저해음

* $t^h\epsilon$ (1.29) } * $s\epsilon$ (1.04) } * $c^h\epsilon$ (0.67)* $t^h\text{I}$ (4.36) } * $s\text{I}$ (0.59) } * $c^h\text{I}$ (0)* t^hi (2.88) } * si (0.59) } * c^hi (0.01)

순/연구개저해음

* $p\epsilon$ (1.20) } * $p^h\epsilon$ (0.80) * $k^h\epsilon$ (1.18) } * $k\epsilon$ (0.34)* $p\text{I}$ (1.08) } * $p^h\text{I}$ (0.92) * $k^h\text{I}$ (1.45) } * $k\text{I}$ (0.07)* pi (1.15) } * p^hi (0.85) * k^hi (1.38) } * ki (0.13)

MaxEnt문법은 제약서열을 가정하지 않기 때문에 제약들 사이에 위치하는 부등호는 가중치의 크기 순을 표시하는 것일 뿐, 제약서열을 의미하지는 않는다. (9)의 가중치 제약문법은 최종적으로 <표 4>와 같은 확률분포를 도출한다. 이 확률분포는 학습데이터의 빈도분포와 매우 유사하지만 정확히 일치하지는 않는다. 3.1절에서 언급했듯이, MaxEnt문법이 학습데이터의 빈도분포에 근접한 확률분포들 가운데 학습데이터로 알 수 있는 정보에 대해서는 그 최대치를 반영하고 학습데이터에 나타나지 않은 불필요한 정보는 최대한도로 배제한 확률분포를 도출하기 때문이다. 이 확률분포를 기저형의 확률까지 포함하여 시각화하면 <그림 4>, <그림 5>와 같다.

표 4. 한국어 체언말 저해음 변이의 확률분포

(%)	c^h -말음 체언			t^h -말음 체언				p^h -말음 체언				k^h -말음 체언			
	-ε	-I	-i	-ε	-I	-i		-ε	-I	-i		-ε	-I	-i	
[c ^h]	65.26	73.47	74.94	33.24	69.48	65.57									
[s]	29.23	26.38	24.4	16.74	28.07	24.01									
[t ^h]	5.51	0.14	0.66	50.01	2.47	10.42									
								[p ^h]	65.43	59.77	69.39				
								[p]	34.57	40.22	36.61				
												[k ^h]	14.8	9.2	10.33
												[k]	85.2	90.8	88.67

<그림 4>는 *cʰ*-말음 체인과 *tʰ*-말음 체인의 기저형-변이형쌍의 확률분포를 누적세로막대형차트(stacked column chart)로 시각화한 것이다. { }로 표시된 차트 제목은 입력형을 나타낸다. 예를 들면, {ch-ε}는 -ε 모음 앞 *cʰ*-말음 체언이다. 가로축은 변이형을 나타낸다. 따라서 [chɛ]는 -ε 모음 앞에서 실현되는 [cʰ]-변이형을 의미한다. 세로축은 확률분포이다. 각 세로막대 그래프는 기저형에 따라 분할되어 있다. 김정색은 /cʰ/-기저형으로 도출된 변이형, 진한 회색은 /s/-기저형으로부터 도출된 변이형, 옅은 회색은 /tʰ/-기저형으로부터 도출된 변이형을 나타낸다.

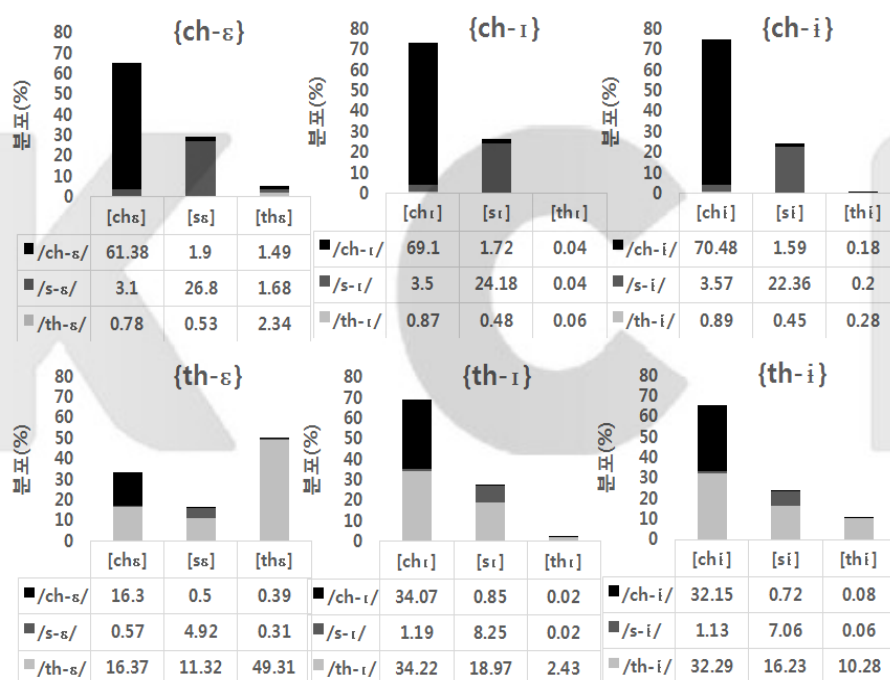


그림 4. 설정저해음 변이의 확률분포

예를 들면, {ch-ε} 차트에서 [chɛ] ([cʰ]-변이형, [cʰɛ])의 확률은 65.26%이다. 이것은 -ε 모음 앞에서 /cʰ/-기저형→[cʰ]-변이형 (/cʰ-ε/→[cʰɛ])이 발생할 확률 61.38%, /s/-기저형→[cʰ]-변이형 (/s-ε/→[cʰɛ])이 발생할 확률 3.1%, /tʰ/-기저형→[cʰ]-변이형 (/tʰ-ε/→[cʰɛ])이 발생할 확률 0.78%를 더한 결과이다.

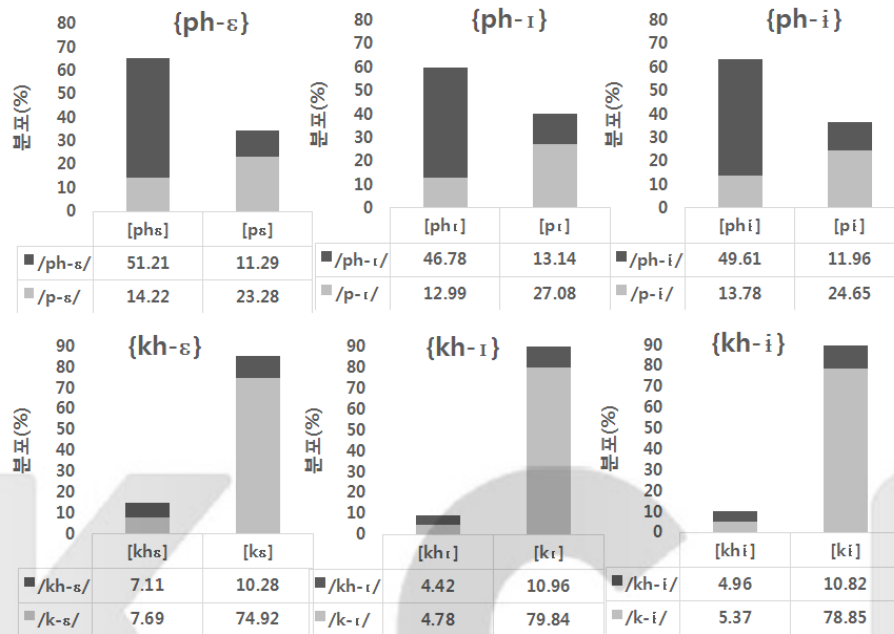


그림 5. 순/연구개저해음 변이의 확률분포

<그림 5>는 p^h -말음 체언과 k^h -말음 체언의 기저형-변이형쌍의 확률분포를 동일한 방식으로 시각화한 것이다: 세로막대 그래프에서 검정색은 /p^h, k^h/-기저형으로부터 도출된 변이형, 옅은 회색은 /p, k/-기저형으로부터 도출된 변이형을 나타낸다. <그림 4>와 <그림 5>에서 제시되듯이, 이 문법의 특징은 도출된 확률분포가 학습데이터의 빈도분포에서 겹으로 드러나지 않는 복수기저형에서 도출되는 변이형의 조건부 확률들이 포함되어 있다는 것이다. 다시 말해서, 이 문법은 입력형 x 와 변이형 y 에 대해 복수의 기저형 z_1, z_2, \dots, z_n 이 매개가 되는 $x-z_1-y, x-z_2-y, \dots, x-z_n-y$ 의 실현확률을 도출한다. 이 도출은 MaxEnt 문법에 복수기저형 체계를 적용시킨 결과이다.

4.2 분석과 토론

이 연구에 따르면, 동일모음 앞에서 나타나는 c^h -말음 체언과 l^h -말음 체언의 확률분포 차이는 UR제약의 종류와 그들 사이의 가중치의 차이로

인해 발생한다. 이것은 복수기저형 체계를 택할지라도 말음을 기준으로 한 체언 유형의 구분이 이 두 체언 유형 사이의 변이형 분포패턴의 차이에 결정적인 영향을 끼침을 보여 준다. 충실성제약은 기저형과 변이형 사이의 연결관계를 평가하는 제약인데 변이형 모두가 기저형이 될 수 있으므로, 충실성제약은 이 차이에 아무런 영향을 끼치지 않는다. 유표성제약은 체언 유형과 관계없이 모음 앞 변이형의 유표성만을 평가하므로, 유표성 제약 역시 동일모음 앞에서 나타나는 이 두 체언 유형 사이의 변이형 분포패턴의 차이에 아무런 영향을 끼치지 않는다. 다시 말해서, 이 두 체언 유형 사이의 변이형 분포패턴의 차이는 c^h -말음 체언 관련 UR제약 $\{c^h\} \rightarrow /c^h/$ (2.66) $\{c^h\} \rightarrow /s/$ (2.20) $\{c^h\} \rightarrow /t^h/$ (0.01)과 t^h -말음 체언 관련 UR제약 $\{t^h\} \rightarrow /t^h/$ (2.58) $\{t^h\} \rightarrow /c^h/$ (0.86) $\{t^h\} \rightarrow /s/$ (0.03) 사이의 가중치 차이로 인해 발생한다.

한편, 동일유형의 체언에서 후행모음에 따른 변이형 분포패턴의 차이는 유표성제약들 사이의 가중치 차이로 인해 발생한다. 동일유형의 체언에서는 후행모음과 관계없이 UR제약과 충실성제약이 동일하게 적용되므로, 유표성제약만이 후행모음에 따른 변이형 분포패턴의 차이에 영향을 끼친다. (9)의 제약 가중치에 따르면, $[c^h]$ -변이형의 실현을 금하는 유표성제약들의 가중치와 $[s]$ -변이형의 실현을 금하는 유표성제약들의 가중치는 크지 않고, 후행모음에 따른 가중치 차이 역시 크지 않다. 반면에, $[t^h]$ -변이형의 실현을 금하는 유표성제약들은 가중치가 클 뿐만 아니라 후행모음에 따른 차이 또한 크다: $*t^h_i$ (4.36), $*t^h_i$ (2.88), $*t^h_e$ (1.29). 따라서 $*t^h_i$ (4.36), $*t^h_i$ (2.88), $*t^h_e$ (1.29) 사이에서 나타나는 가중치 차이가 후행모음에 따른 변이형 분포패턴의 차이에 큰 영향을 끼친다. 충실성제약은 UR제약, 유표성제약과 상호작용하면서 동일모음 앞 각 체언 유형의 변이형들의 발생확률을 결정하는데 역할을 한다.

지금부터는 $-e$ 모음 앞 c^h -말음 체언의 변이형 확률분포와 $-e$ 모음 앞 t^h -말음 체언의 변이형 확률분포, $-i$ 모음 앞 t^h -말음 체언의 변이형 확률분포에 대한 보다 상세한 분석을 통해 이 연구의 결과에 대해 토론한다. 먼저, 제약들을 가중치 순으로 배열한 <표 5>를 가지고 $-e$ 모음 앞 c^h -말음 체언의 변이형들의 확률분포를 살펴본다.

표 5. c^h -말음 체언 + $-ε$ 모음⁵

기저형	변이형	F- c^h 3.11	{ c^h }→/ c^h / 2.66	F- s 2.52	{ c^h }→/ s / 2.20	F- t^h 1.72	* $t^hε$ 1.29
/c ^h -ε/	[c ^h ε]				-1		
/s-ε/	[c ^h ε]		-1	-1			
/t ^h -ε/	[c ^h ε]		-1		-1	-1	
/c ^h -ε/	[sε]	-1			-1		
/s-ε/	[sε]		-1				
/t ^h -ε/	[sε]		-1		-1	-1	
/c ^h -ε/	[t ^h ε]	-1			-1		-1
/s-ε/	[t ^h ε]		-1	-1			-1
/t ^h -ε/	[t ^h ε]		-1		-1		-1

기저형	변이형	*sε 1.04	*c ^h ε 0.67	{ c^h }→/ t^h / 0.01	H	e^H	p
/c ^h -ε/	[c ^h ε]		-1	-1	-2.88	0.0561	61.41
/s-ε/	[c ^h ε]		-1	-1	-5.86	0.0029	3.12
/t ^h -ε/	[c ^h ε]		-1		-7.25	0.0007	0.78
/c ^h -ε/	[sε]	-1		-1	-6.36	0.0017	1.89
/s-ε/	[sε]	-1		-1	-3.71	0.0245	26.78
/t ^h -ε/	[sε]	-1			-7.62	0.0005	0.54
/c ^h -ε/	[t ^h ε]			-1	-6.61	0.0013	1.74
/s-ε/	[t ^h ε]			-1	-6.48	0.0015	1.68
/t ^h -ε/	[t ^h ε]				-6.15	0.0021	2.33

<표 5>에서 [c^hε], [sε], [t^hε] 각각은 세 기저형 /c^h-ε/, /s-ε/, /t^h-ε/ 모두로부터 도출된다. 기저형-변이형쌍의 H-value를 비교하면 변이형 [c^hε], [sε], [t^hε] 사이의 확률 차이가 UR제약과 유표성제약으로 인해 생기고, 각 변이형 내의 기저형에 따른 확률 차이는 UR제약과 충실성제약으로 인해 생겼음을 알 수 있다. <표 5>에 따르면, c^h -말음 체언의 기저형으로는 /c^h-/기저형이 선호되고 변이형으로는 [c^h]-변이형이 선호된다.

먼저, 기저형이 그대로 실현된 /c^h-ε/→[c^hε] 61.41%, /s-ε/→[sε] 26.78%, /t^h-ε/→[t^hε] 2.33% 사이의 확률 차이가 어떻게 발생했는지를 살펴보자. 이

⁵ <표 5>의 결과 가운데 일부는 소수점 아래 자리의 수치가 <그림 4>와 다르다. 이것은 <그림 4>가 R에서 제약 가중치의 소수점 아래 자리 수치를 반올림하지 않은 채 확률분포를 계산한 결과인데 반해 <표 5>는 수작업으로 소수점 셋째 자리에서 반올림하여 확률분포를 계산하였기 때문에 발생한 것이다.

경우 충실성제약은 언제나 만족되므로 충실성제약은 변수가 될 수 없다. 이 차이는 UR제약 $\{c^h\} \rightarrow /c^h/$ (2.66), $\{c^h\} \rightarrow /s/$ (2.20), $\{c^h\} \rightarrow /t^h/$ (0.01)의 가중치 차이와 유표성제약 $*[c^h\epsilon]$ (0.67), $*[s\epsilon]$ (1.04), $*[t^h\epsilon]$ (1.29)의 가중치 차이가 결합된 결과이다. $/c^h-\epsilon/ \rightarrow [c^h\epsilon]$ 와 $/s-\epsilon/ \rightarrow [s\epsilon]$ 사이의 H-value의 차이는 0.83인데 이것은 $\{c^h\} \rightarrow /c^h/$ (2.66)의 가중치와 $*[s\epsilon]$ (1.04)의 가중치를 더한 값에서 $\{c^h\} \rightarrow /s/$ (2.20)의 가중치와 $*[c^h\epsilon]$ (0.67)의 가중치를 더한 값을 뺀 값과 동일하다: $(2.66+1.04) - (2.20+0.67)$. 그리고 $/c^h-\epsilon/ \rightarrow [c^h\epsilon]$ 와 $/t^h-\epsilon/ \rightarrow [t^h\epsilon]$ 사이의 H-value의 차이는 3.27인데, 이것은 $\{c^h\} \rightarrow /c^h/$ (2.66)의 가중치와 $*[t^h\epsilon]$ (1.29)의 가중치를 더한 값에서 $\{c^h\} \rightarrow /t^h/$ (0.01)의 가중치와 $*[c^h\epsilon]$ (0.67)의 가중치를 더한 값을 뺀 값과 동일하다: $(2.66+1.29) - (0.01+0.67)$.

이제, $[c^h]$ -변이형의 경우를 예로 들어 기저형에 따른 변이형 내의 확률 차이를 살펴보자. $[c^h]$ -변이형이 실현될 경우에는 기저형과 관계없이 유표성제약 $*c^h\epsilon$ (0.67)를 위반한다. 따라서 유표성제약은 변이형 내의 기저형에 따른 확률 차이를 초래하는 변수가 될 수 없다. 이 차이는 UR제약과 충실성제약의 결과이다. $/c^h-\epsilon/ \rightarrow [c^h\epsilon]$ 는 가중치가 큰 UR제약 $\{c^h\} \rightarrow /c^h/$ (2.66)를 만족시키고 이보다 가중치가 작은 두 UR제약 $\{c^h\} \rightarrow /s/$ (2.20)과 $\{c^h\} \rightarrow /t^h/$ (0.01)를 위반한다. 유표성제약 $*c^h\epsilon$ (0.67)도 위반하므로 $/c^h-\epsilon/ \rightarrow [c^h\epsilon]$ 의 H-value는 -2.88이다. $/s-\epsilon/ \rightarrow [c^h\epsilon]$ 는 UR제약 $\{c^h\} \rightarrow /s/$ (2.20)를 만족시키는 반면, 다른 UR제약들 $\{c^h\} \rightarrow /c^h/$ (2.66), $\{c^h\} \rightarrow /t^h/$ (0.01)와 충실성제약 Faith-s (2.52)를 위반한다. 따라서 H-value는 $/c^h-\epsilon/ \rightarrow [c^h\epsilon]$ 보다 훨씬 작아진다 (H-value -5.86). 한편, $/t^h-\epsilon/ \rightarrow [c^h\epsilon]$ 는 가중치가 매우 작은 UR제약 $\{c^h\} \rightarrow /t^h/$ (0.01)를 만족시키는 대신, 가중치가 큰 두 UR제약 $\{c^h\} \rightarrow /c^h/$ (2.66)와 $\{c^h\} \rightarrow /s/$ (2.20)를 위반하고 충실성제약 Faith- t^h (1.72) 역시 위반하므로 H-value가 현저히 작아진다 (H-value -7.25). 따라서 이 분석은 c^h -말음 체언의 $/c^h/$ -기저형이 $[c^h]$ -변이형으로 실현되는 확률이 매우 높은 이유를 적절하게 설명한다.

$[s]$ -변이형의 경우, $/c^h-\epsilon/ \rightarrow [s\epsilon]$ 는 가중치가 큰 UR제약 $\{c^h\} \rightarrow /c^h/$ (2.66)을 만족시킨다. 그러나 이 제약을 위반하는 대신 가중치가 작은 UR제약 $\{c^h\} \rightarrow /s/$ (2.20)을 만족시키는 $/s-\epsilon/ \rightarrow [s\epsilon]$ (26.78%)보다 훨씬 확률이 낮다 ($/c^h-\epsilon/ \rightarrow [s\epsilon]$ (1.89%)). 이것은 $/c^h-\epsilon/ \rightarrow [s\epsilon]$ 가 가중치가 큰 충실성제약 Faith- c^h (3.11)를 위반하는 반면, $/s-\epsilon/ \rightarrow [s\epsilon]$ 는 Faith- c^h 를 만족시키기 때문에 생긴 결과이다. 따라서 이 분석은 c^h -말음 체언의 $[s]$ -변이형의 확률은 주로 화자들 가운데 일부가 c^h -말음 체언의 기저형으로 $/s/$ -기저형을 택한 결과임을 보여 준다.

<표 4>에 따르면, -i와 -i 모음 앞 c^h -말음 체언의 변이형 분포패턴 역시 지금까지 살펴본 -ε 모음 앞 변이형 분포패턴과 유사하다. 그러나 $[t^h]$ -변이형의 확률이 -ε 모음 앞에서보다 -i와 -i 모음 앞에서 훨씬 낮다는 점은 논

의할 만한 가치가 있다: [tʰɛ]가 5.51%인데 반해, [tʰɪ]는 0.14%, [tʰi]는 0.66%이다. *cʰ*-말음 체언과 관련된 UR제약과 충실성제약은 후행모음과 관계없이 공통적으로 적용되므로, 이 차이는 후행모음에 따른 유표성제약들 사이의 가중치 차이의 결과이다. 후행모음과 관계없이 유표성제약이 선호하는 변이형의 순서는 [cʰ]-변이형 > [s]-변이형 > [tʰ]-변이형이지만, 그 가중치는 후행모음에 따라 달라진다. 즉, *tʰɛ (1.29) > *sɛ (1.04) > *cʰɛ (0.67)과 *tʰɪ (4.36) > *sɪ (0.59) > *cʰɪ (0), 그리고 *tʰi (2.88) > *si (0.59) > *ci (0.01)에서 보듯이, -ɪ와 -i 모음 앞에서 [tʰ]의 실현을 금하는 *tʰɪ (4.36)와 *tʰi (2.88)의 가중치가 -ɛ 모음 앞에서 [tʰ]의 실현을 금하는 *tʰɛ (1.29)의 가중치보다 크다. 결국, 이 차이가 [tʰɛ]와 [tʰɪ]/[tʰi] 사이의 확률 차이를 유발한다.

지금까지의 분석에 따르면, *cʰ*-말음 체언의 경우 기저형 선택은 변이형 확률분포에 큰 영향을 끼친다. 각 변이형들은 동일형태의 기저형으로부터 도출되는 경우가 대부분인데, 이것은 *cʰ*-말음 체언의 경우 유표성제약이 선호하는 변이형의 순서가 UR제약과 충실성제약의 상호작용이 선호하는 변이형의 순서와 서로 상충하지 않기 때문에 생긴 결과이다. 그러나 이 분석은 또한 기저형 선택이 UR제약들에 의해서만 결정되는 것이 아니라 UR제약, 충실성제약, 유표성제약의 상호작용 즉, 언어공동체에 내재된 전반적인 문법에 의해 결정된다는 것을 보여 준다. 더욱이 [tʰɛ]와 [tʰɪ], [tʰi] 사이의 확률 차이가 유표성제약들로부터 기인한다는 사실로부터 기저형 선택만이 변이형의 확률분포에 영향을 끼치는 유일한 요인이 아님을 알 수 있다.

이제, 지금까지 살펴본 -ɛ 모음 앞 *cʰ*-말음 체언 변이의 확률분포를 아래 <표 6>에 제시된 -ɛ 모음 앞 *tʰ*-말음 체언 변이의 확률분포와 비교해 보자.

표 6. *tʰ*-말음 체언 + -ɛ모음

기저형	변이형	F- <i>cʰ</i> 3.11	{tʰ} → /tʰ/ 2.58	F- <i>s</i> 2.52	F- <i>tʰ</i> 1.72	*tʰɛ 1.29	*sɛ 1.04
/cʰ-ɛ/	[cʰɛ]		-1				
/s-ɛ/	[cʰɛ]		-1	-1			
/tʰ-ɛ/	[cʰɛ]				-1		
/cʰ-ɛ/	[sɛ]	-1	-1				-1
/s-ɛ/	[sɛ]		-1				-1
/tʰ-ɛ/	[sɛ]				-1		-1
/cʰ-ɛ/	[tʰɛ]	-1	-1			-1	
/s-ɛ/	[tʰɛ]		-1	-1		-1	
/tʰ-ɛ/	[tʰɛ]					-1	

기저형	변이형	$\{t^h\} \rightarrow /c^h/$ 0.86	$*c^h\epsilon$ 0.67	$\{t^h\} \rightarrow /s/$ 0.03	H	e^H	p
$/c^h-\epsilon/$	$[c^h\epsilon]$		-1	-1	-3.28	0.0376	16.38
$/s-\epsilon/$	$[c^h\epsilon]$	-1	-1		-6.63	0.0013	0.57
$/t^h-\epsilon/$	$[c^h\epsilon]$	-1	-1	-1	-3.28	0.0376	16.38
$/c^h-\epsilon/$	$[s\epsilon]$			-1	-6.76	0.0012	0.5
$/s-\epsilon/$	$[s\epsilon]$	-1			-4.48	0.0113	4.93
$/t^h-\epsilon/$	$[s\epsilon]$	-1		-1	-3.65	0.0260	11.31
$/c^h-\epsilon/$	$[t^h\epsilon]$			-1	-7.01	0.0009	0.39
$/s-\epsilon/$	$[t^h\epsilon]$	-1			-7.25	0.0007	0.31
$/t^h-\epsilon/$	$[t^h\epsilon]$	-1		-1	-2.18	0.1130	49.21

UR제약들이 t^h -말음 체언과 관련된 것들로 대체되었을 뿐, 나머지 제약들의 종류와 가중치는 <표 5>와 동일하다. c^h -말음 체언 관련 UR제약들에서는 $\{c^h\} \rightarrow /c^h/$ (2.66)과 $\{c^h\} \rightarrow /s/$ (2.20)의 가중치가 크고 $\{c^h\} \rightarrow /t^h/$ (0.01)의 가중치가 작았으나, t^h -말음 체언 관련 UR제약들에서는 $\{t^h\} \rightarrow /c^h/$ (0.86), $\{t^h\} \rightarrow /s/$ (0.03)의 가중치가 작은 반면, $\{t^h\} \rightarrow /t^h/$ (2.58)의 가중치가 크다. 체언 유형에 따른 UR제약들의 이러한 가중치 차이는 변이형들의 분포 차이 즉, $[c^h\epsilon] \succ [s\epsilon] \succ [t^h\epsilon]$ 대 $[t^h\epsilon] \succ [c^h\epsilon] \succ [s\epsilon]$ 로 나타난다. 그러나 $-\epsilon$ 모음 앞 t^h -말음 체언 변이형들 사이의 확률 차이가 c^h -말음 체언의 경우에 비해 크지 않다는 점은 논의를 필요로 한다.

주목할 만한 점은 <표 6>에 따르면 t^h -말음 체언의 경우 $/c^h/$ -기저형과 $/s/$ -기저형으로부터 도출된 $[t^h]$ -변이형의 확률은 낮은 반면, $/t^h/$ -기저형으로부터 도출된 $[c^h]$ -변이형과 $[s]$ -변이형의 확률은 상대적으로 높다는 사실이다. 이와 같은 결과는 t^h -말음 체언에 적용되는 UR제약들의 가중치와 관련되어 있다. $/c^h/$ -기저형으로부터 $[t^h]$ -변이형이 도출된 확률은 0.39%에 불과하다. t^h -말음 체언이 $/c^h/$ -기저형을 택함으로 인해 가중치가 큰 UR제약 $\{t^h\} \rightarrow /t^h/$ (2.58)를 위반한 것이 이 낮은 확률에 결정적인 영향을 끼친다. $/s/$ -기저형으로부터 $[t^h]$ -변이형이 도출된 확률 역시 마찬가지이다. 이 경우의 확률은 0.31%인데, UR제약 $\{t^h\} \rightarrow /t^h/$ (2.58)의 위반이 결정적 영향을 끼친다. 반면에, $/t^h/$ -기저형으로부터 $[c^h]$ -변이형이 도출된 확률은 16.38%로 $/c^h/$ -기저형으로부터 도출된 경우의 확률과 거의 비슷하다⁶. 이것은 $/t^h/$ -기저형을 택할 경우 가중치가 작은 UR제약 $\{t^h\} \rightarrow /c^h/$ (0.86)을 위반하고 가중치가 큰

⁶ <표 6>에서 이 둘은 16.38%로 동일하다. 그러나 <그림 4>에서는 16.37%와 16.3%로 서로 다르다. 이 차이 역시 각주 (5)에서 기술한 이유로 인해 초래된 것이다.

UR제약 $\{t^h\} \rightarrow /t^h/$ (2.58)를 만족시키기 때문에 생긴 결과이다. 한편, $/t^h/$ -기저형으로부터 $[s]$ -변이형이 도출된 확률은 11.31%로 다른 기저형들로부터 $[s]$ -변이형이 도출된 확률보다 훨씬 높다: $/c^h/$ -기저형으로부터 도출된 확률은 0.5%이고 $/s/$ -기저형으로부터 도출된 확률은 4.93%이다. 이것 역시 $/t^h/$ -기저형을 택할 경우 가중치가 작은 UR제약 $\{t^h\} \rightarrow /s/$ (0.03)을 위반하지만, 가중치가 큰 UR제약 $\{t^h\} \rightarrow /t^h/$ (2.58)를 만족시키기 때문에 생긴 결과이다.

이러한 결과는 기저형 선택의 패턴에 있어서 t^h -말음 체언은 앞에서 이미 토론한 c^h -말음 체언과 매우 다르다는 것을 보여 준다. c^h -말음 체언의 경우에는 각 변이형이 동일형태의 기저형으로부터 도출되는 경우가 대부분이었고 기저형 선택이 변이형 확률분포에 끼치는 영향은 매우 컸다. 그러나 <표 6>에 따르면, t^h -말음 체언의 경우 $[s]$ -변이형은 형태가 다른 기저형으로부터 도출되는 경우가 대부분이고, $[c^h]$ -변이형은 기저형 선택의 영향에서 비교적 자유롭다. 이것은 t^h -말음 체언의 경우에도 유표성제약과 충실성제약은 $[c^h]$ -변이형 > $[s]$ -변이형 > $[t^h]$ -변이형 순으로 선호하지만 t^h -말음 체언 관련 UR제약들이 $/t^h/$ -기저형을 매우 선호하기 때문에 생긴 결과이다.

여기에서 제기되어야 할 중요한 질문은 c^h -말음 체언 관련 UR제약들의 가중치와 t^h -말음 체언 관련 UR제약들의 가중치가 왜 다른가이다. 이에 대한 답은 학습데이터의 전반적 변이형 빈도분포에서 찾아야 한다. 학습데이터에 나타난 c^h -말음 체언의 변이형 분포패턴은 후행모음과 관계없이 일정하지만, t^h -말음 체언의 변이형 분포패턴은 후행모음에 따라 큰 차이를 보인다. t^h -말음 체언이 $/t^h/$ -기저형을 택할 것을 요구하는 UR제약 $\{t^h\} \rightarrow /t^h/$ (2.58)의 큰 가중치는 학습데이터에 나타난 $-e$ 모음 앞 $[t^h]$ -변이형의 높은 빈도 즉, 48.4%를 반영한 결과이다. 만약 $-i$ 와 $-i$ 모음 앞에서처럼, $-e$ 모음 앞에서도 $[t^h]$ -변이형의 빈도가 낮고 $[c^h]$ -변이형의 빈도가 높았다면, UR제약 $\{t^h\} \rightarrow /t^h/$ 의 가중치는 그렇게 크지 않았을 것이다. 한편, 학습데이터의 빈도분포에 따르면, t^h -말음 체언은 $-i$ 와 $-i$ 모음 앞에서 $[c^h]$ -변이형으로 자주 실현된다. 그러나 이 경우의 높은 빈도에도 불구하고 t^h -말음 체언이 $/c^h/$ -기저형을 택할 것을 요구하는 UR제약 $\{t^h\} \rightarrow /c^h/$ (0.86)의 가중치는 매우 작다. 대신 이 빈도는 $-i$ 와 $-i$ 모음 앞에서 $[t^h]$ -변이형의 실현을 금하는 유표성제약 $*t^h_i$ (4.36), $*t^h_i$ (2.88)의 큰 가중치와 $[c^h]$ -변이형의 실현을 금하는 유표성제약 $*c^h_i$ (0), $*c^h_i$ (0.01)의 작은 가중치에 반영되어 있다.

지금까지 살펴본 $-e$ 모음 앞 t^h -말음 체언의 변이형 확률분포를 아래 <표 7>에 제시된 $-i$ 모음 앞 t^h -말음 체언의 변이형 확률분포와 비교해 보자.

표 7. t^h -말음 체언 + -ɪ 모음

기저형	변이형	* t^h_I 4.36	F- c^h 3.11	{ t^h }→/ t^h / 2.58	F- s 2.52	F- t^h 1.72	{ t^h }→/ c^h / 0.86
/c ^h -ɪ/	[c ^h ɪ]			-1			
/s-ɪ/	[c ^h ɪ]			-1	-1		-1
/t ^h -ɪ/	[c ^h ɪ]					-1	-1
/c ^h -ɪ/	[sɪ]		-1	-1			
/s-ɪ/	[sɪ]			-1			-1
/t ^h -ɪ/	[sɪ]					-1	-1
/c ^h -ɪ/	[t ^h ɪ]	-1	-1	-1			
/s-ɪ/	[t ^h ɪ]	-1		-1	-1		-1
/t ^h -ɪ/	[t ^h ɪ]	-1					-1

기저형	변이형	* s_I 0.59	{ t^h }→/ s / 0.03	* c^h_I 0	H	e^H	p
/c ^h -ɪ/	[c ^h ɪ]		-1	-1	-2.61	0.0735	34.15
/s-ɪ/	[c ^h ɪ]			-1	-5.96	0.0026	1.2
/t ^h -ɪ/	[c ^h ɪ]		-1	-1	-2.61	0.0735	34.15
/c ^h -ɪ/	[sɪ]	-1	-1		-6.31	0.0018	0.84
/s-ɪ/	[sɪ]	-1			-4.03	0.0178	8.25
/t ^h -ɪ/	[sɪ]	-1	-1		-3.2	0.0408	18.93
/c ^h -ɪ/	[t ^h ɪ]		-1		-10.08	4e-05	0.02
/s-ɪ/	[t ^h ɪ]				-10.32	3e-05	0.02
/t ^h -ɪ/	[t ^h ɪ]		-1		-5.25	0.0052	2.44

<표 6>과 <표 7>의 비교에서 주된 관심은 -ɛ 모음 앞에서 t^h -말음 체언이 [t^h]-변이형이 될 확률과 [c^h]-변이형이 될 확률은 각각 50.01%, 33.24%인데 반해, -ɪ 모음 앞에서 이 변이형들의 확률은 각각 2.47%와 69.48%로 크게 다르다는 점이다. <표 6>과 <표 7>은 유표성제약들만 다를 뿐, 나머지 제약들의 종류와 가중치는 동일하다. 따라서 이러한 차이를 유발하는 요인은 유표성제약이다. <표 6>에서 유표성제약들은 * t^h_ϵ (1.29) } * s_ϵ (1.04) } * c^h_ϵ (0.67)이었으나, <표 7>에서는 * t^h_I (4.36) } * s_I (0.59) } * c^h_I (0)이다. 즉, -ɪ 모음 앞 t^h 의 실현을 금하는 * t^h_I (4.36)의 가중치는 -ɛ 모음 앞 t^h 의 실현을 금하는 * t^h_ϵ (1.29)의 가중치보다 매우 크다. 그러나 -ɪ 모음 앞 c^h 와 s 의 실현을 금하는 * c^h_I (0)와 * s_I (0.59)의 가중치는 -ɛ 모음 앞 c^h 와 s 의 실현을 금하는 * c^h_ϵ (0.67)와 * s_ϵ (1.04)의 가중치보다 작다. * t^h_I (4.36)의 가중치가 매우 크기 때문에, 이를 위반한 /c^h-ɪ/→[t^hɪ], /s-ɪ/→[t^hɪ], /t^h-ɪ/→[t^hɪ]의 H-value는 * t^h_ϵ (1.29)

를 위반한 / c^h - ϵ ->[$t^h\epsilon$], / s - ϵ ->[$t^h\epsilon$], / t^h - ϵ ->[$t^h\epsilon$]의 H-value보다 훨씬 작아진다. 그 결과, t^h -말음 체언의 -i 모음 앞 [t^h]-변이형의 확률이 - ϵ 모음 앞 [t^h]-변이형의 확률보다 훨씬 낮아진다. 특히, / t^h - ϵ ->[$t^h\epsilon$] (49.21%)의 H-value는 -2.18이고, / t^h -i->[t^hi] (2.44%)의 H-value는 -5.25로 큰 차이를 보이는데, 이 차이가 결정적 영향을 끼친다. $*c^h\epsilon$ (0.67)와 $*s\epsilon$ (1.04)에 비해 작은 $*c^hi$ (0)와 $*si$ (0.59)의 가중치로 인해, [c^h]-변이형과 [s]-변이형의 -i 모음 앞 확률은 - ϵ 모음 앞 확률보다 높아진다.

이 분석은 t^h -말음 체언의 후행모음에 따른 변이형 분포패턴의 차이가 유표성제약들 사이의 가중치 차이의 결과임을 보여 준다. c^h -말음 체언의 경우에는 유표성제약처럼 UR제약과 충실성제약의 상호작용 역시 [t^h]-변이형보다는 [c^h]-변이형을 더 선호하기 때문에 후행모음에 따른 유표성제약들 사이의 가중치 차이가 변이형 분포패턴에 큰 영향을 끼치지 않는다. 따라서 c^h -말음 체언의 경우에는 후행모음과 관계없이 변이형의 분포패턴이 일정하다.

후행모음과 관련된 유표성제약들 사이의 가중치 차이가 t^h -말음 체언의 후행모음에 따른 변이형 분포패턴에 결정적인 영향을 끼친다는 사실은 t^h -말음 체언의 변이가 전통적 음운분석에서 제안되었던 / t^h ->[c^h]/___i 구개음화의 영향에서 자유롭지 않음을 보여 준다. 그리고 이 연구에서 구개음화의 영향은 $*t^hi$ (4.36)와 $*t^h\epsilon$ (1.29)와 같이 가중치 차이라는 계량적 값의 차이로 형식화되었다. -i 모음 앞에서도 [c^h]-변이형의 확률이 높고 [t^h]-변이형의 확률은 낮는데, 이 또한 -i 모음과 관련된 유표성제약들 $*t^hi$ (2.88) > $*si$ (0.59) > $*c^hi$ (0.01)의 가중치 차이에 의한 결과이다. 따라서 이 연구의 분석은 t^h -말음 체언의 변이가 구개음화 환경이 -i 모음 앞에서 -i, -i 모음 앞으로 확장되는 과정에서 나타난 결과라고 제안한 Han (2002), Kim (2005), Sohn (2012) 등의 선행연구들을 뒷받침한다. 이 분석은 또한 변이형 확률분포를 예측하기 위해서는 구개음화가 / t^h ->[c^h]/___i, i와 같이 범주적으로 형식화되기보다는 가중치 제약들을 가지고 형식화되는 변이문법이 필요하다는 것을 암시한다.

순/연구개저해음-말음 체언의 변이패턴은 t^h -말음 체언보다는 c^h -말음 체언의 변이패턴과 더 유사하다. c^h -말음 체언처럼, UR제약, 충실성제약, 유표성제약들의 요구가 서로 상충하지 않고, 유표성제약들 사이의 가중치 차이도 후행모음에 따라 크게 달라지지 않는다. 이 체언들의 변이패턴 분석에서 주된 관심은 p^h -말음 체언은 전통적 기저형이 그대로 실현되는 변이형 ([p^h]-변이형)을 선호하는데 반해, k^h -말음 체언은 대체형 ([k]-변이형)을 선호한다는 점이다. c^h -말음 체언처럼, 이 두 유형의 체언에서도 각 변이형은

주로 동일형태의 기저형으로부터 도출된다. 앞의 <그림 5>에 따르면, -ε 모음 앞 [p^h]-변이형의 확률은 65.43%인데, 이 확률은 /p^h/-기저형으로부터 도출된 경우의 확률 51.21%와 /p/-기저형으로부터 도출된 경우의 확률 14.22%를 더한 결과이다. 그리고 -ε 모음 앞 [p]-변이형의 확률 34.57%는 /p/-기저형으로부터 도출된 경우의 확률 23.28%와 /p^h/-기저형으로부터 도출된 경우의 확률 11.29%를 더한 결과이다. 이와 같은 경향은 -i와 -i 모음 앞에서도 동일하게 관찰된다. k^h-말음 체언에서는 -ε 모음 앞 [k^h]-변이형의 확률은 14.8%인데 반해, 대체형인 [k]-변이형의 확률은 85.2%이다. [k]-변이형의 확률 85.2%는 /k/-기저형으로부터 도출된 경우의 확률 74.92%와 /k^h/-기저형으로부터 도출된 경우의 확률 10.28%를 더한 결과로, [k]-변이형은 주로 /k/-기저형으로부터 도출된다. 이와 같은 경향은 -i와 -i 모음 앞에서도 동일하게 관찰된다. 따라서 p^h-말음 체언과 k^h-말음 체언 사이의 변이형 분포패턴의 차이는 주로 기저형 선택의 차이에 기인한 것임을 알 수 있다.

지금까지의 토론에서 알 수 있듯이, 이 연구는 복수기저형 가중치 제약 문법을 채택하여 한국어 체언말 저해음 변이를 분석하였다. 이 문법은 기저형-변이형쌍의 확률을 제시함으로써 학습데이터만 가지고 겉으로는 알 수 없는 화자들의 기저형 선택에 관한 감추어진 지식을 형식화하였다. 이 문법의 적용 결과에 따르면, 한국어 체언말 저해음 변이는 아래 <그림 6>과 같이 크게 세 가지의 유형으로 구분될 수 있다: 이 세 가지 유형은 범주적으로 구분되는 것이 아니라 점진적/확률적으로 구분된다.

한국어 체언말 저해음 변이는 기저형 선택이 결정적인 영향을 끼치는 변이와 음운과정이 결정적인 영향을 끼치는 변이로 구분된다. 지금까지의 분석에 따르면, 변이패턴에서 보이는 이러한 차이에 결정적인 영향을 끼치는 것은 구개음화로부터 얼마나 자유로울 수 있는가이다. c^h-말음 체언의 경우에는 UR제약과 충실성제약 사이의 상호작용의 요구가 구개음화를 선호하는 유표성제약의 요구와 상충되지 않는다. p^h-말음 체언과 k^h-말음 체언은 구개음화와 무관하다. 결국, c^h-말음 체언, p^h-말음 체언, k^h-말음 체언들의 변이형 분포패턴은 상당 부분 화자들이 서로 다른 기저형을 선택하거나 한 화자가 복수의 기저형을 선택하기 때문에 생긴 결과이다. 한편, l^h-말음 체언은 구개음화와 관련된 음운과정이 결정적인 영향을 끼치는 반면, 기저형 선택이 변이형 확률분포에 끼치는 영향은 상대적으로 약하다.

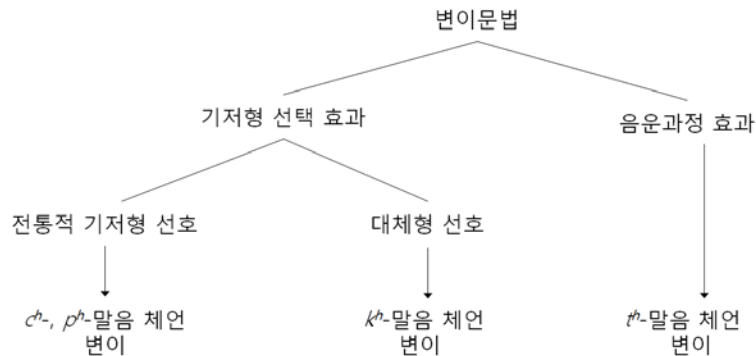


그림 6. 한국어 체언말 저해음 변이의 분류

아직 토론되지 않은 중요한 질문은 변이형들의 분포패턴과 단어들의 사용빈도 또는 단어들의 어휘부 분포패턴 사이의 관계이다. 이 연구가 채택한 복수기저형 MaxEnt문법에서 각 변이형은 복수기저형에서 도출된다⁷. 따라서 이 질문은 단어들의 사용빈도 또는 단어들의 어휘부 분포패턴이 기저형 선택에 어떤 영향을 끼치는지에 대한 질문과 관련된다. 이 연구는 기저형에 대한 혼동이 일어나거나 기억이 불확실하여 화자들이 대체기저형을 찾을 때 대체기저형 선택에 영향을 끼치는 것이 바로 단어들의 사용빈도 또는 단어들의 어휘부 분포패턴이라고 제안한다. Jun (2010: 156)은 기본형(base form)에 대한 규칙의 적용과 관련하여 이와 유사한 제안을 하였다.

강범모·김흥규 (2004)의 말뭉치를 조사한 Jun and Lee (2007: 165)에 따르면, 설정저해음-말음 체언 총 822개 가운데 s-말음 체언 52.07% (428개), cʰ-말음 체언 27.62% (227개), tʰ-말음 체언 17.15% (141개), t-말음 체언 0.85% (7개), c-말음 체언 2.31% (19개)이다. 그리고 순저해음-말음 체언 총 2,451개 가운데 p-말음 체언 96.7% (2,370개), pʰ-말음 체언 3.3% (81개)이고, 연구개음-말음 체언 총 7,305개 가운데, k-말음 체언은 99.58% (7,274개), kʰ-말음 체언은 0.3% (22개), k'-말음 체언 0.12% (9)개이다. 따라서 기저형에 대한 혼동이 생길 때 가장 빈도가 높은 /s/-기저형, /p/-기저형, /k/-기저형을 대체가능한 기저형으로 선택하는 것이라고 추론할 수 있다. 한편, cʰ-말음 체언에서 /tʰ/-기저형을 선택가능한 기저형으로 삼고, tʰ-말음 체언에서 /cʰ/-기저형을

⁷ 설정저해음-말음 체언에서는 [c]-변이형과 [t]-변이형도 나타난다. 이들의 빈도가 매우 낮기 때문에 이 연구에서는 무시하고 다루지 않았지만, 원칙적으로는 이들과 동일형태의 기저형들도 선택가능한 기저형들로 취급되어야 한다.

선택가능한 기저형으로 삼는 것은 구개음화로 인해 l^h -말음 체언이 $[c^h]$ -변이형으로 실현되는 경우가 빈번해서 이 두 기저형에 대한 혼동이 발생한 결과라고 볼 수 있다.

그러나 이 연구는 단어들의 사용빈도 또는 단어들의 어휘부 분포패턴이 변이형들의 분포패턴에 끼치는 영향은 제한적이라고 제안한다. MaxEnt문법이 발화실험 결과를 학습하여 가중치 제약문법을 구축하듯이, 대체형을 기저형으로 선택한 화자들은 변이데이터의 변이형 빈도분포에 기반하여 가중치 제약문법을 구축한다. 이와 같은 기저형 선택이나 음운작용의 영향없이 단어들의 사용빈도 또는 단어들의 어휘부 분포패턴이 변이형 확률분포를 전적으로 결정한다는 가정 하에서는 왜 c^h -말음 체언과 l^h -말음 체언에서 $[s]$ -변이형이 가장 높은 확률로 실현되지 않는지에 대한 적절한 설명을 제공하기 어렵다. 특히, l^h -말음 체언에서 $[s]$ -변이형보다 $[c^h]$ -변이형의 확률이 더 높게 나타나는 현상에 대한 타당한 이유를 제시할 수 없을 것이다. 순저해음-말음 체언 변이와 연구개음-말음 체언 변이의 경우도 마찬가지다. k -말음 체언과 l^h -말음 체언의 어휘부 빈도관계처럼, p -말음 체언도 p^h -말음 체언보다 압도적으로 더 높은 어휘부 빈도를 나타낸다. 그럼에도 불구하고, 왜 p^h -말음 체언의 경우에는 $[p^h]$ -변이형의 확률이 $[p]$ -변이형의 확률보다 더 높은지에 대한 적절한 설명을 제공할 수 없다.

5. 선행연구와 비교

5.1 데이터 비교

선행연구와의 비교는 크게 두 부분으로 나뉜다. 이 절에서는 데이터 수집과 관련된 점을 비교하고, 5.2절에서는 이론적 분석을 중심으로 비교한다. 여러 선행연구들은 발음이 한글철자로 적힌 설문지를 이용해 참여자들에게 단어가 어떻게 발음되는지(최혜원 2004, Sohn 2012), 발음 적절성 점수가 얼마인지(Kang 2003), 어떤 발음을 선호하는지(강은지 외 2004)를 묻거나, 한글철자로 적힌 단어를 발화하도록 하는(Jun and Lee 2007) 방식 등을 통해 데이터를 수집하였다. 이에 반해, 이 연구는 그림을 이용하여 수행한 발화실험의 결과를 데이터로 삼았다. 여기에서는 선행연구들의 데이터 가운데 여러 연구들에서 인용된 바 있는 최혜원(2004)과 강은지 외(2004)의 데이터를 이 연구의 발화실험 결과와 비교한다.

최혜원(2004)은 40대~70대 한국어 화자 총 1,174명을 대상으로 평소 해당 단어를 어떻게 발음하는지를 묻는 선택형 설문지 언어 실태조사를 실

시하였다. 설문지는 c^h -말음 체언, t^h -말음 체언, p^h -말음 체언, k^h -말음 체언 각각 11개, 15개, 10개, 2개, 총 38개의 체언이 -ε, -ɪ, -i, -uy (의), -a (아) 앞에서 어떻게 발음되는지를 묻는 질문들을 포함하고 있다. 선택항목은 <전통적 기저형 그대로([c^h 또는 t^h, p^h, k^h])>, < c^h -말음 체언, t^h -말음 체언일 경우 [s]>, < p^h -말음 체언일 경우 [p]>, < k^h -말음 체언일 경우 [k]>, <잘 모르겠음>, <기타>로 구성되어 있다. -i 모음 앞 t^h -말음 체언의 경우에는 선택항목으로 [c^h]가 추가되었고, -a 모음 앞 t^h -말음 체언과 c^h -말음 체언의 경우에는 [t]가 추가되었다. 후행모음이 -uy, -a일 경우를 제외한 결과는 <표 8>과 같다.

표 8. 후행모음에 따른 변이형의 빈도분포 (최혜원 2004)

(%)	c^h -말음 체언			t^h -말음 체언			p^h -말음 체언			k^h -말음 체언		
	-ε	-ɪ	-i	-ε	-ɪ	-i	-ε	-ɪ	-i	-ε	-ɪ	-i
[c ^h]	48.5	61.9	55.7	---	60.8	24.3	[p ^h] [p] [k ^h] [k]	63.8 36.2 41.6 58.4	63.2 36.8 35.9 64.1	65.3 34.7 27.3 72.7		
[s]	51.5	38.1	44.3	24	39.2	24						
[t ^h]	---	---	---	76	---	50.8						

최혜원(2004)에서는 각 체언별 빈도가 제시되어 있다. 그러나 체언 유형별 빈도는 전통적 기저형으로 실현된 경우와 -ɪ 모음 앞에서 t^h 가 [c^h]로 실현된 경우 이외에는 제시되어 있지 않다. 특히, <잘 모르겠음>과 <기타>를 선택한 빈도가 제시되어 있지 않기 때문에, <표 8>에서 굵은 글씨 부분만이 최혜원(2004)으로부터 알 수 있는 결과이고, 나머지 부분은 <잘 모르겠음>과 <기타>를 선택한 빈도를 무시하고 산출된 것이므로 최혜원(2004)의 실제 조사 결과보다 다소 높은 추정치이다. 이 결과가 발화수행이 아닌 실태조사로부터 나온 것이고 한글철자로 제시된 선택항목 가운데 하나를 택하도록 하여 나온 결과이므로, 발화실험 결과와 단순비교하는 데에는 무리가 따른다. 따라서 여기에서는 전반적인 분포패턴만을 비교할 것이다.

c^h -말음 체언의 -ε 모음 앞 [c^h]-변이형의 빈도는 우리의 발화실험 결과보다 더 낮고 [s]-변이형의 빈도는 더 높다. 반면, 발화실험 결과와 마찬가지로 -ɪ와 -i 모음 앞에서는 [c^h]-변이형의 빈도가 [s]-변이형의 빈도보다 더 높다. t^h -말음 체언의 -ε와 -ɪ 모음 앞 변이형 분포패턴은 우리의 발화실험 결과와 유사하지만, -i 모음 앞 분포패턴은 매우 다르다. 즉, 발화실험 결과

처럼, -ε 모음 앞에서는 [tʰ]-변이형의 빈도가 [s]-변이형의 빈도보다 높고, -i 모음 앞에서는 [cʰ]-변이형의 빈도가 [s]-변이형의 빈도보다 높다. 그러나 -i 모음 앞에서는 발화실험 결과와 달리, [tʰ]-변이형의 빈도가 가장 높다. pʰ-말음 체언과 kʰ-말음 체언의 각 변이형의 구체적인 빈도는 우리의 발화실험 결과와 다르지만, 분포패턴은 유사하다. 즉, 후행모음과 관계없이 pʰ-말음 체언에서는 [pʰ]-변이형, kʰ-말음 체언에서는 [k]-변이형의 빈도가 더 높다.

요약하면, cʰ-말음 체언의 -ε 모음 앞 [s]-변이형의 빈도가 [cʰ]-변이형의 빈도보다 더 높다는 점과 tʰ-말음 체언의 -i 모음 앞 [tʰ]-변이형의 빈도가 다른 변이형들보다 더 높다는 점을 제외하면, 전반적인 분포패턴은 우리의 발화실험 결과와 유사하다. 특히, 최혜원(2004)의 조사결과에서도 cʰ-말음 체언, pʰ-말음 체언, kʰ-말음 체언은 후행모음과 관계없이 변이형의 분포패턴이 대체로 일정한데 반해, tʰ-말음 체언은 후행모음에 따라 변이형의 분포패턴이 심하게 달라진다. 따라서 이 결과 역시 체언을 유형별로 구분하지 않고는 한국어 체언말 저해음의 변이패턴을 설명할 수 없음을 보여 준다. 또한, tʰ-말음 체언의 경우 -i 모음 앞에서 [tʰ]-변이형의 빈도가 가장 높게 나타나지만 [cʰ]-변이형의 빈도 역시 24.3%에 이른다는 것은 발화실험 결과와 마찬가지로 구개음화 음운환경이 -i 모음 앞으로까지 확장되는 과정에 있음을 보여 준다. 앞에서 밝혔듯이, 각 변이형의 빈도가 발화실험 결과와 약간씩 차이가 나는 것은 조사방법의 차이 즉, 설문지 언어 실태조사 대 발화실험 간 차이 또는 한글철자의 영향 여부가 작용한 것으로 추론할 수 있다.

최혜원(2004)과 달리, 강은지 외(2004)는 어떻게 단어를 발음하는지를 묻는 대신 설문참여자도 선호하는 발음을 선택하게 하는 선택형 설문지를 사용하는 조사를 실시하였다. 참여자는 20대~70대 한국어 화자 총 156명이었다. 설문지는 s-말음 체언, c-말음 체언, cʰ-말음 체언, tʰ-말음 체언, pʰ-말음 체언, kʰ-말음 체언 각각 3개, 4개, 5개, 4개, 5개, 2개, 총 23개의 체언이 -ε, -i, -i 모음 앞에서 어떻게 발음되는 것을 선호하는지를 묻는 질문을 포함하고 있다. 이밖에도 체언말 자음연쇄 등에 대한 질문도 포함하고 있다.

<표 9>는 s-말음 체언과 c-말음 체언의 경우를 제외한 결과를 제시한 것이다. 빈도가 적혀 있지 않은 항목은 이들의 설문에 해당 모음환경에서 해당 저해음이 발화될 경우가 포함되어 있지 않아 조사가 이루어지지 않은 경우이다. 이 조사의 결과가 보이는 특징은 전반적으로 체언유형과 관계없이 한글철자에 나타난 전통적 기저형의 빈도가 우리의 발화실험 결과보다 높다는 것이다. 그러나 주목할 만한 것은 이 조사에서도 tʰ-말음 체언의 경우에는 후행모음에 따라 변이형의 분포패턴이 달라진다.

표 9. 후행모음에 따른 변이형의 빈도분포 (강은지 외 2004)

(%)	c^h -말음 체언			t^h -말음 체언						
	-ε	-I	-i	-ε	-I	-i				
[c ^h]	---	79.1	80.6	1.4	---	33.2				
[s]	---	20.9	19.4	1.9	---	9.1				
[t ^h]	---	0	0	95.8	---	57.7				
							p^h -말음 체언			
							-ε	-I	-i	
							[p ^h]	76.9	37.1	86.4
							[p]	23.1	62.9	13.6
							k^h -말음 체언			
							-ε	-I	-i	
							[k ^h]	47.7	---	68.2
							[k]	52.3	---	31.8

우리의 발화실험 결과와 달리, 최혜원(2004)과 강은지 외(2004)에서 공통적으로 나타나는 현상은 μ -말음 체언의 경우 전통적 구개음화 환경이 아닌 - ϵ 모음과 - i 모음 앞에서 [ɲ]-변이형이 매우 높은 빈도를 나타낸다는 것이다. 발화실험에서 그 빈도는 각각 48.4%, 10.5%였는데 반해, 최혜원에서는 76%, 50.8%, 강은지 외에서는 95.8%, 57.7%이다. 발화실험이 아닌 상황에서는 한글철자에 나타난 체언 유형의 영향으로 구개음화 적용환경의 확대가 억제되지만, 한글철자가 제시되지 않은 발화상황에서는 그 영향이 상대적으로 약화되어 구개음화 적용환경이 상대적으로 덜 제한되었기 때문에 나타난 것으로 이 차이를 해석할 수도 있을 것이다⁸.

5.2 이론적 분석 비교

변이를 예외로 보지 않고 의미있는 분석대상으로 본 연구들 가운데 상대적으로 이른 시기에 이루어진 연구들은 주로 범주적 관점에 입각하여 이 변이를 다루었다. 이들은 주로 복수의 변이형들을 도출하는 문법작용을 밝히는데 초점을 맞췄다. 예를 들면, Han (2002)은 변이형의 출현을 패러다임균일성(paradigm uniformity) 효과에 기인한 것으로 보았으며, Kim (2005)은

8 심사자 가운데 한 분은 이 차이를 실험/조사참여자의 연령 차이에 기인한 것일 수 있음을 제안하였다. 즉, 최혜원(2004)과 강은지 외(2004)의 조사대상자가 우리의 발화 실험참여자들의 연령대보다 상대적으로 높기 때문에 발생한 것일 수 있음을 지적하였다. 그러나 강은지 외(2004)의 20-30대 참여자들의 *l*^h-말음 체인의 -ε와 -i 모음 앞 [ɱ]-변이형의 빈도 역시 매우 높은 것으로 보아(각각 96.41%, 56.75%), 이 차이가 연령 차이에 기인한 것은 아닌 것으로 보인다.

패러다임 균일성 효과와 기저형 재구조화 효과 사이의 상호작용의 결과라고 보았다.

고광모 (1989)와 Kang (2003)은 변이형의 분포패턴과 단어들의 사용빈도 또는 단어들의 어휘부 분포패턴 사이의 관계에 주목했다는 점에서 범주적 접근과 구분된다. 고광모(1989)는 [s]-변이형, [p]-변이형, [k]-변이형들이 빈번하게 나타나는 것은 이 저해음들로 끝나는 한국어 체언이 매우 많기 때문에 발생하는 유추(analogy)의 결과라고 보았다. 한편, 앞서 살핀 최혜원 (2004)과 강은지 외 (2004)처럼, Kang (2003)은 설문지를 사용한 실태조사를 수행하였다. 참여자는 8명이었고 c-말음 체언, cʰ-말음 체언, t-말음 체언 각각 3개, 5개, 8개가 대상 체언이었으며 후행모음 환경은 -ε, -i, -i였다. 최혜원 (2004)과 강은지 외 (2004)와 다른 것은 Kang (2003)은 [s]-변이형에만 초점을 맞췄다는 점이다. 각 체언에 대해 [s]-변이형을 제시하고 적절성(goodness)을 1(나쁨)~4(좋음)의 점수로 평가하도록 하였다. 조사결과에 따르면, 체언 사용빈도가 낮을수록 [s]-변이형의 적절성은 더 높았다. ‘음성적 요인에 의한 변화(phonetically motivated change)’는 고빈도 단어에서 먼저 발생하고 ‘유추적 변화(analogically motivated change)’는 저빈도 단어에서 먼저 발생한다는 가정 하에, Kang (2003) 역시 설정저해음-말음 체언에서 [s]-변이형의 출현은 유추적 변화의 결과라고 제안하였다. 이 연구들은 전반적인 변이형 분포패턴, 이를 테면, cʰ-말음 체언과 t-말음 체언 사이의 분포패턴의 차이, t-말음 체언에서 후행모음에 따른 분포패턴의 차이, pʰ-말음 체언과 kʰ-말음 체언 사이의 선호 변이형 차이를 형식화하는 데까지 연구 범위를 확장하지는 않았다.

Jun and Lee (2007), Albright (2008), Jun (2010)은 연폐쇄음 t, p, k를 기본형(base form)으로 삼아 변이형 분포패턴의 형식화를 시도했다는 점에서 앞서 살핀 연구들과 구분된다. 여기에서는 Jun (2010)만을 살펴본다. Jun (2010)은 체언말 설정저해음의 변이형 빈도 (또는 선호도)는 [s] > [cʰ, tʰ] > [c, t]인데, 이 순서가 어휘부 체언말 저해음의 빈도분포 또는 사용빈도의 순서와 동일하다는 점에 주목하였다. 그리고 [cʰi]와 [cʰε] 사이, [tʰi]와 [tʰε] 사이의 빈도는 각각 [cʰi] > [cʰε], [tʰε] > [tʰi]인데, 이 순서 역시 어휘부 체언말 빈도분포 또는 사용빈도의 순서와 동일하다는 점 역시 주목하였다. Jun (2010)은 이러한 동일성은 연폐쇄음 [t]를 모든 설정저해음의 기본형으로 삼고 이 기본형을 다른 저해음으로 실현시키는 규칙, 이를 테면, t → sil / X__# (기본형이 [t]로 끝나면, 해당 대격형은 [sil]로 끝나야 한다), t → cʰil / X__# (기본형이 [t]로 끝나면, 해당 대격형은 [cʰil]로 끝나야 한다)와 같은 규칙들(Jun 2010: 158)의 조정신뢰도(adjusted confidence) 크기 순서의 결과라고

제안하였다. 이 조정신뢰도는 말뭉치에서 해당 모음 앞에 나타나는 저해음 전체 (범위(scope))에 대한 해당 저해음 (적중치(hit))의 비율이다. 예를 들면, 말뭉치에서 -il 앞에 나타나는 전체 저해음의 수가 291개이고, sil의 수는 157개, c^hil의 수는 68개이므로, $t \rightarrow \text{sil} / X_ \#$ 의 원신뢰도(raw confidence)는 0.540, 조정신뢰도(adjusted confidence)는 0.520이고 $t \rightarrow \text{c}^{\text{h}}\text{il} / X_ \#$ 의 원신뢰도는 0.234, 조정신뢰도 0.218이다⁹. 그의 제안에 따르면, 기본형은 해당 체언의 기저형이 아닌 연폐쇄음 [t]이고 어휘부 체언말 저해음의 빈도 또는 사용빈도에 따라 체언말 저해음의 변이형 분포가 결정된다.

Jun (2010)의 제안은 한국어 체언말 저해음 변이의 전반적 분포패턴을 적절히 예측하였다는 점에서 큰 의미를 지닌다. 그러나 체언의 유형을 구분하지 않고 변이형의 빈도를 산출하였기 때문에, 모음을 기준으로 한 각 유형별 체언의 변이형 분포패턴의 차이를 어떻게 설명할 것인지에 대한 과제를 남긴다. 특히, 후행모음에 따라 달라지는 분포패턴을 보이는 ^h-말음 체언 변이형들의 특성을 구개음화의 영향을 배제한 채 어떻게 설명할 수 있을지에 대한 질문을 남긴다.

Jun (2010)과는 달리, 체언의 유형을 구분할 것을 제안했다는 점에서 이 연구의 제안과 유사한 Sohn (2012)은 [s]-변이형만이 어휘부의 체언말 /s/의 빈도에 영향을 받는다고 보고 이를 ‘유추적 평준화(analogical leveling)’라고 하였다: “어간말 설정저해음은 어휘부의 /s/로 끝나는 어간의 높은 빈도를 반영하여 전체 패러다임에 걸쳐 연마찰음 /s/로 평준화된다(the stem-final coronal obstruents are leveled to the lax fricative /s/ across the paradigm, reflecting the relatively dense population of the /s/-final stem in the lexicon) (Sohn 2012: 58-59).” Sohn (2012)에 따르면, 나머지 변이형들의 출현은 구개음화 적용 환경의 확대와 재분석(reanalysis)의 결과이다. 변이형의 빈도 $/t^{\text{h}}\text{l}/ \rightarrow [c^{\text{h}}\text{l}] > /t^{\text{h}}\text{i}/ \rightarrow [c^{\text{h}}\text{i}] > /t^{\text{h}}\text{e}/ \rightarrow [c^{\text{h}}\text{e}]$ 는 구개음화 환경이 ‘-l 모음 앞’에서 ‘-l와 -i 모음 앞’으로 확대되는 과정에서 패러다임 평준화가 일어난 결과이고, $/c^{\text{h}}\text{e}/ \rightarrow [t^{\text{h}}\text{e}] > /c^{\text{h}}\text{i}/ \rightarrow [t^{\text{h}}\text{i}] > /c^{\text{h}}\text{l}/ \rightarrow [t^{\text{h}}\text{l}]$ 는 $/c^{\text{h}}\text{e}/$ 를 $/t^{\text{h}}\text{e}/$ 로 재분석한 결과이다.

⁹ Mikheev(1997)를 따라 Jun은 작은 사례수에 기초해 만들어진 규칙에 대해서 별점을 부여하기 위해, 신뢰하한값 통계를 사용한다. 원신뢰도는 범위에 대한 적중치의 비율을 구하지만, 조정신뢰도는 $\pi = \hat{\rho}^* - t_{(1-\alpha)/2}^{(n-1)} * S = \hat{\rho}^* - t_{(1-\alpha)/2}^{(n-1)} * \sqrt{\frac{\hat{\rho}^*(1-\hat{\rho}^*)}{n}}$ 식을 가지고 구한다(Mikheev 1997: 413). 이 때 $\hat{\rho}_i^* = \frac{x_i + 0.5}{n_i + 1}$ 이며, $t_{(1-\alpha)/2}^{(n-1)}$ 는 t -분포의 계수(coefficient)이다.

6. 결론

이 연구는 한 언어공동체의 화자들이 한 형태소/단어에 대해 서로 다른 기저형을 선택하는 것이 음운변이를 일으키는 중요한 요인일 수 있다는 가정으로부터 시작되었고 이 가정을 확인하기 위해 발화실험을 수행하였다. 그리고 Pater et al. (2012)이 제안한 복수기저형 MaxEnt문법을 채택하여 복수기저형 체계를 형식화하는 동시에, 발화실험 결과를 학습데이터로 삼아 한국어 체언말 저해음 변이를 위한 가중치 제약 확률문법을 제안하였다. 이 문법은 각 변이형에 대한 복수기저형들의 확률분포를 포함하고 있는 변이형 확률분포를 도출하였다. 이 확률분포로부터 체언의 유형 구분을 고려하지 않고는 한국어 체언말 저해음 변이를 설명하기 어려움을 확인하였다. 그리고 한국어 체언말 저해음 변이가 체언의 유형에 따라 구분되는 세 가지 다른 변이 유형의 복합현상임을 제안하였다. 발화실험 결과와 제약의 가중치, 변이형들의 확률분포에 따르면, c^h -말음 체언과 p^h -말음 체언은 전통적으로 기저형이라고 여겨진 것과 동일형태의 변이형을 더 선호하고, k^h -말음 체언은 대체형을 더 선호하는 반면, t^h -말음 체언의 변이패턴은 구개음화의 영향 때문에 기저형 선택이 끼치는 영향은 상대적으로 약하였다.

기저형-변이형쌍의 확률분포가 UR제약, 충실성제약, 유표성제약의 상호작용의 결과라는 사실은 기저형 선택이 한국어 공동체에 내재된 전반적인 문법에 의해 결정된다는 것을 보여 주었다. 따라서 이 전반적인 문법을 구성하는 단어간 변이문법, 개별 화자들에게 투사될 때 발현되는 개별문법, 그리고 개별문법과 관련된 화자들의 군집화(clustering)를 분석할 필요가 있었다. 그러나 이 연구에서는 이를 전혀 다루지 않았다. 이에 대한 분석이 이루어질 때 한국어 체언말 저해음 변이의 특성은 명확하게 드러날 것이다. 따라서 이 연구는 이와 관련한 후속 연구를 과제로 남긴다.

참고문헌

- 강범모·김홍규. 2004. *한국어 형태소 및 어휘 사용 빈도의 분석 2*. 서울: 고려 대학교 민족 문화 연구원.
- 강은지, 이호영, 김주원. 2004. 서울말 어간말 자음의 음성 실현. *말소리* 49, 1-30.
- 고광모. 1989. 체언 끝의 변화 ㄷ>ㅅ에 대한 새로운 해석. *언어학* 11, 3-22.
- 유재원. 1985. *우리말 역순 사전*. 서울: 정음사.

- 최혜원. 2004. *표준 발음 실태 조사 3*. 서울: 국립국어원.
- 홍성훈. 2014. *음운론의 계량적 방법론*. 서울: 한국문화사.
- ALBRIGHT, ADAM. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Language Universals and Language Change*, 144–181. Oxford: Oxford University Press.
- GOLDWATER, SHARON and MARK JOHNSON. 2003. Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111-120.
- GUY, GREGORY R. 2003. Variationist approaches to phonological change. In Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, 369-400.
- HAN, EUNJO. 2002. Optimal paradigms in Korean nominal inflection. *Studies in Phonetics, Phonology and Morphology* 8.2, 303–322. The Phonology-Morphology Circle of Korea.
- JÄGER, GERHARD. 2007. Maximum entropy models and stochastic Optimality Theory. In Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson and Annie Zaenen (eds.), *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*, 467-479. Stanford: CSLI Publications.
- JOHNSON, MARK. 2002. Optimality-theoretic lexical functional grammar. In Suzanne Stevenson and Paolo Merlo (eds.), *The Lexical Basis of Sentence Processing: Formal, Computational and Experimental Issues* 4, 59-73.
- JUN, JONGHO. 2010. Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19.2, 137-179.
- JUN, JONGHO and JEEHYUN LEE. 2007. Multiple stem-final variants in Korean native nouns and loanwords. *Journal of the Linguistic Society of Korea* 47, 159-187.
- KANG, YOONJUNG. 2003. Sound changes affecting noun-final coronal obstruents in Korean. *Japanese/Korean Linguistics* 12, 128-139.
- KIM, JIN-HYUNG. 2005. A reconsideration of phonological leveling: A case of noun inflection in Korean. *Studies in Phonetics, Phonology and Morphology* 11.2, 259–274. The Phonology-Morphology Circle of Korea.
- MIKHEEV, ANDREI. 1997. Automatic rule induction for unknown-word guessing. *Computational Linguistics* 23, 405-423.
- PATER, JOE, KAREN JESNEY, ROBERT STAUBS and BRIAN SMITH. 2012. Learning probabilities over underlying representations. *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and*

Phonology, 62-71.

R DEVELOPMENT CORE TEAM. 2015. R: A Language and Environment for Statistical Computing (Version 3.2.3) [Computer program]. <http://www.R-project.org>.

SHANNON, CLAUDE E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27.3, 379-423, 623-656.

SOHN, HYANG-SOOK. 2012. Random patterning in paradigms of noun-final coronal obstruents. *Studies in Phonetics, Phonology and Morphology* 18.1, 51-84. The Phonology-Morphology Circle of Korea.

STAUBS, ROBERT. 2011. *Tableau.R/solver.R/solvertest.R*. [R-code]. <http://blogs.umass.edu/hgr/hg-in-r> (with Joe Pater, Karen Jesney and Brian Smith).

WILSON, COLIN. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30.5, 945-982.

김선희

06974 서울특별시 동작구 흑석로 84
중앙대학교 인문대학 영어영문학과
e-mail: sunhoi@cau.ac.kr

김하나

06974 서울특별시 동작구 흑석로 84
중앙대학교 인문대학 영어영문학과
e-mail: hanang11@daum.net

received: June 30, 2016

revised: August 5, 2016

accepted: August 16, 2016