

2017 한국음운론학회 여름학술대회
2017. 6. 17(토) 서울시립대학교

한국어 명사의 음소배열제약

-통계적 학습과 단어성 판단 조사-

서울대학교 언어학과 박사과정
박나영 (arimnet@naver.com)



음소배열제약의 정의

- 음소배열제약: 음소 단위의 결합 또는 회피에 대한 적형성 인식
- 예: 영어의 음소배열제약 *_w[lb]

✓ 비단어 lbick
☞ 제약을 위배, 비적형으로 인식됨

cf. 비단어 blick
☞ 제약을 위배하지 않음, 적형으로 인식됨

- ‘적형-비적형’의 이분법적 인식 → 범주적인 음소배열제약 포착

비범주적 적형성 인식

- 다수 연구가 비범주적 적형성 인식을 보고
 - Vitevitch et al. (1997), Albright (2006), Shademan (2007)
- 실험을 통해, 영어 '[비음]-[저해음]'에 대한 세분화된 적형성 인식을 밝힘(Hay et al. 2003)
 - ✓ **적형성** /nt/ > /nf/ > /mk/ > /ms/
- 어휘부(lexicon)의 빈도가 적형성 인식을 결정한다고 주장
 - ✓ **고빈도** /nt/
 - ✓ **중빈도** /nf/ > /mk/ > /ms/

- 비범주적 적형성과 빈도와의 관계를 통계적 문법으로 포착하려는 시도
 - ✓ Coleman & Pierrehumbert (1997)
 - Stochastic Context-free grammar
(통계적 문맥 자유 문법)
 - ✓ Hayes & Wilson (2008)
 - A Maximum Entropy Model of Phonotactics
(최대 엔트로피 음소배열제약 모델)

한국어의 음소배열제약

➤ 빈도 0인 연쇄에 대한 음소배열제약

- ✓ 제약 예1: 한국어 음절 두음에 [ŋ]가 허용되지 않는다.

*
 $\sigma \left[\begin{array}{l} +nasal \\ +dorsal \end{array} \right]$ 즉, * $\sigma[\eta]$

- ✓ 제약 예2: 설정 저해음과 /j/계 이중모음의 연속을 허용하지 않는다.

*
 $\left[\begin{array}{l} -sonorant \\ +coronal \end{array} \right] \left[\begin{array}{l} -back \\ -syllabic \end{array} \right]$

- ✓ 제약 예3: 저해음과 공명음의 연쇄를 허용하지 않는다.

*
 $\left[-sonorant \right] \left[+sonorant \right]$

➤ 각 제약은 음성학적, 음운론적으로 자연스러움

비범주적 적형성 인식에 대한 단서

- 저빈도 연쇄에 대한 직관적 서술(허웅 1985)
 - ✓ 예: [i]은 w-계 이중모음, [je, jɛ] 결합이 잘 나타나지 않는다.

- 계량적 연구
 - ✓ 음소 연쇄에 대한 빈도 조사
 - ✓ 연쇄 결합 관계를 자질로 표시한 제약으로 제시
 - 유재원(1997): 후설모음은 장애음과의 결합을 꺼린다.

- 계량적 연구는 비범주적 적형성 인식을 체계적으로 포착하지 못함
 - ✓ 해당 연쇄가 제약으로 기능할 수 있는 발생 빈도에 대해 기준이 분명하지 않음
 - ✓ 양적 정보가 문법에 직접 반영될 수 있는 기제가 없음

통계적 문법 도입

- 연쇄의 양적 정보가 문법에 직접 반영된다고 가정
 - ✓ 음소 결합 관계를 통계에 기초하여 파악하고, 그 실재를 점검
 - [자음]-[모음], [모음]-[자음]의 발생 빈도와 이에 근거한 통계치를 계산하고, 이에 따른 선호도 인식을 밝힘 (이용은 2009)
- Cho (2012): 최대 엔트로피 음소배열제약 모델 채택
 - ✓ 어휘부 내에서 발생 빈도가 낮은 연쇄를 제약으로 학습
 - 예: $^* \begin{bmatrix} -high \\ -back \end{bmatrix} [+lateral]$ 즉, $^*[ei]$ 가중치: 3.772
 - ✓ 개별 연쇄에 대한 비범주적 적형성을 예측할 수 있음

이제까지의 조사 대상: 한국어 전체

- 대부분의 양적 연구가 어휘부를 구분하지 않고 조사
- Cho (2012)의 학습 어휘부
 - ✓ 차용어를 제외한 5,702 단어 학습(조남호 2003)
 - ✓ 품사 및 어원 구분을 하지 않음
 - ✓ ‘한국어 전체’를 학습자의 어휘부로 가정

어휘부가 세분화될 가능성 제기

- 단어의 어원에 따라, 음소 출현 빈도가 다를 수 있음이 언급됨
 - 예: **한자어**는 **고유어**와 비교해 ‘ㄱ, ㄴ’가 많이 사용되고, ‘ㅡ’가 상대적으로 적게 사용된다. (안소진 2009)
- 단어의 형태론적 구성이 음소 결합의 회피 조건이 될 수 있음
 - 예: 현대 국어의 **형태소 내부**에서는 ‘ㄹ’ 뒤에 오는 설정 저해음은 평음이 될 수 없고, 격음 또는 경음이어야 한다. (고광모 1996)

연구 목적

- 세분화된 어휘부에 대한 문법 학습
- 개별 어휘부 문법이 적형성 판단에 영향을 미치는지 살피고자 함
 - ✓ 개별 어휘부 문법 간 상대적 기여도 점검

연구 방법

문법: 최대엔트로피 음소배열제약 모델

대상: 한국어 명사 및 세분화된 어휘부

단어성 판단 조사(1-7점)

2음절어 조사

단음절어 조사

문법 예측과 실제 적형성 인식 비교

평균

상관관계

혼합효과 선형모델

최대 엔트로피 음소배열제약 모델 (HAYES & WILSON 2008)

- 가정: 음소 연쇄의 발생 빈도 \approx 음소 연쇄에 대한 적형성
- 입력: 어휘부 + 자질 목록 \rightarrow 출력: 제약 + 가중치
- 제약의 형태 (*[자질] - [자질], 가중치)
 - ✓ 유표성 제약만이 학습된다.
 - ✓ 가중치
 - 해당 연쇄의 회피 정도를 나타냄

- 해당 연쇄의 비적형성 점수(harmony)
 - ✓ 제약의 가중치를 통해서, 비적형성 점수를 계산할 수 있음
 - ✓ 해당 연쇄의 적형성은 하나 이상의 제약을 위배할 때, 위배되는 제약들의 가중치 합, 즉 비적형성 점수로 예측됨
 - ❖ 따라서, 비적형성 점수가 높을수록 비적형으로 인식될 것을 예측
- 모델의 특징
 - ✓ 음소배열제약에 대한 귀납적 모델
 - ✓ 음소의 결합 관계를 체계적으로 살필 수 있음
 - ✓ 어휘부의 양적 정보를 문법에 직접 반영할 수 있는 기제

어휘부의 구성

- 한국어 화자의 적형성 인식이 ‘명사’에서 비롯된다고 가정
 - ✓ 명사는 한국어 품사의 대다수를 차지
 - ✓ 한국어 화자들이 새로운 단어를 대부분 ‘명사’로 인지
 - ✓ 한자어와 고유어에 대한 충분한 자료를 보장받기 위함

- 명사: 37,157 단어
 - ✓ 빈도 5 이상인 일반명사(강범모·김흥규 2009)
 - ✓ 표준국어대사전(<http://www.korean.go.kr>)에 등재된 단어

학습 대상 어휘부: 단어

- **학습 어휘부 1: 명사 전체** 예: 강바닥, 규산염
 - ✓ 한국어 연쇄에 대한 대부분의 계량적 연구는 어휘부를 구분하지 않음
 - ✓ 본 연구: 명사 전반에 유효한 제약을 탐색하고자 함
- **어원과 단어의 형태론적 구성에 따라, 학습 어휘부를 세분화**

학습 대상 어휘부: 단어

- **학습 어휘부 2: 단일어** (17,980 단어) 예: 하늘, 갈등
- ✓ 일반적으로 음소배열제약 탐색 시, 형태소 1개로 이루어진 단어(즉, 단일어)를 기본 단위로 가정
 - 영어: Hayes&White 2013
- ✓ 한국어의 양적 연구에서는 ‘단일어’에 대한 탐색이 시도되지 않음
- ✓ 본 연구: 단일어 부류를 구성하여, 단일어 음소배열제약을 탐색하고자 함

학습 대상 어휘부: 단어

- **학습 어휘부 3: 고유 단일어** (1,871 단어) 예: 하늘, 가난
 - ✓ 한국어 화자의 순수한 문법 특징이 고유 단일어에 잘 반영되는 것으로 가정됨 (김경일 1985, 한성우 2006, Ito 2007)
- **학습 어휘부 4: 한자 단일어** (16,088 단어) 예: 공기, 갈등
 - ✓ 본 연구: 형태론적 구성 조건을 기준으로, 고유 단일어에 대응하여 구성

분류 기준 및 입력형

➤ 어원

- ✓ 표준국어대사전 원어 정보

➤ 단일어 여부

- ✓ 표준국어대사전 형태소 구분 정보 이용
- ✓ 형태소 구분 정보가 표시되어 있지 않더라도, 단어의 일부가 사전에 등재되어 있다면 해당 단어는 복합어로 분류함
 - 예: 간암

➤ 각 단어는 표준국어대사전 대표 발음형으로 입력

- ✓ 10모음 체계 채택

학습 대상 어휘부: 음절

- 음절을 어휘부의 구성 단위로 가정할 수 있음
 - ✓ 한자어는 어원적으로 구성 음절이 형태소에 대응될 수 있음이 언급됨
 - 한자어의 음소배열제약은 주로 한 음절 내에서 탐색됨
(신지영 2009, 안소진 2009)
 - ✓ 고유어에 대해서도 같은 가정을 적용
- 음절을 어휘부의 구성 단위로 삼아 학습
 - ✓ 고유 단일어와 한자 단일어를 구성하는 기저 음절형 기준
 - ✓ 각 음절과 음절의 유형 빈도를 입력 자료로 사용
- **학습 어휘부 5: 고유 단일어 구성 음절** (679 음절)
- **학습 어휘부 6: 한자 단일어 구성 음절** (450 음절)

학습 시뮬레이션

- 소프트웨어: UCLA phonotactic learner
 - ✓ 입력: 어휘 목록 + 자질 → 출력: 제약 *[자질][자질] + 가중치
 - ✓ 학습 조건: 최대 학습 제약의 수, 제약의 최대 길이, 제약의 정확도

The screenshot shows the UCLA Phonotactic Learner application window. The interface includes a 'Choose a Task' section with four radio button options: 'Start from scratch - discover all new constraints' (selected), 'Test an existing grammar with the test data', 'Add more constraints to an existing grammar', and 'Reweight the constraints of an existing grammar'. A blue box with a left-pointing arrow and the text '새 문법 탐색 선택' is positioned over the first option. Below this is a 'Working Folder' field containing the path 'C:\Users\nayoung\Desktop\whole_2gram_internal' and a 'Change' button. The 'Basic Input Files' section has three rows: 'Features chart' with 'feature0106_final.txt' and a blue box labeled '자질 목록'; 'Projected tiers' with empty fields and 'View'/'Change' buttons; and 'Learning data' with '2whole1Av2.txt' and a blue box labeled '학습 어휘 목록'. The 'Basic Parameters' section has three rows: 'Maximum number of learning constraints' with an empty input field and a blue box labeled '최대 학습 제약의 수'; 'Maximum constraint length' with '2' and a blue box labeled '제약의 최대 길이'; and 'Accuracy' with '0.3' and a blue box labeled '정확도'.

➤ 자질 목록

| | | aspirate | tense | labial | high | low | back | round |
|---|----|-----------------|--------------|---------------|-------------|------------|-------------|--------------|
| ㅍ | p | - | - | + | 0 | 0 | 0 | 0 |
| ㅍ | ph | + | - | + | 0 | 0 | 0 | 0 |
| ㅂ | pp | - | + | + | 0 | 0 | 0 | 0 |
| ㅓ | e | 0 | 0 | 0 | - | - | - | - |
| ㅕ | E | 0 | 0 | 0 | - | + | - | - |
| ㅣ | i | 0 | 0 | 0 | + | - | - | - |
| ㅏ | a | 0 | 0 | 0 | - | + | + | - |

➤ 단어 단위 입력

| <u>명사</u> | | <u>단일어</u> | | <u>한자 단일어</u> | | <u>고유 단일어</u> | |
|-----------|---------------------|------------|-------------|---------------|-----------------|---------------|-----------|
| 강바닥 | k a n g p p a t a k | 공기 | k o n g k i | 공기 | k o n g k i | 하늘 | h a n x l |
| 규산염 | k y u s a n n y v m | 가늌 | k a n x m | 갈등 | k a l t t x n g | 가난 | k a n a n |
| 하늘 | h a n x l | 곤란 | k o l l a n | 가훈 | k a h u n | 가늌 | k a n x m |

➤ 음절 단위 입력 (음절 및 유형 빈도)

| <u>고유 단일어 음절</u> | | | | <u>한자 단일어 음절</u> | | | |
|------------------|-----|----|--|------------------|-----|-----|--|
| 가 | ka | 62 | | 가 | ka | 219 | |
| 각 | kak | 4 | | 각 | kak | 123 | |
| 갈 | kal | 12 | | 갈 | kal | 15 | |

학습 조건

➤ 개별 어휘부별 학습 조건 요약

| | 단어 문법 학습 | 음절 문법 학습 |
|-------------|----------|----------|
| 제약의 최대 길이 | 2 | 3 |
| 최대 학습 제약의 수 | 제한하지 않음 | 150 |
| 정확도 | 0.3 | |

➤ 각 어휘부별로 5번씩 학습

학습 조건

- 단어 문법 학습: 두 음소 결합 관계를 탐색

제약: $*[+labial] \begin{bmatrix} +high \\ +back \\ -round \end{bmatrix}$

예: $*[pɪ]$

- 음절 문법 학습: 세 음소의 결합 관계까지 살핌으로써 회피 음절을 일반화된 문법 제약으로 포착하고자 함

제약: $*[+labial] \begin{bmatrix} -low \\ +back \\ -round \end{bmatrix} \begin{bmatrix} -sonorant \\ +dorsal \end{bmatrix}$

예: $*[pɪk, pʌk]$

학습 결과 요약

➤ 학습 제약의 수

| | | | |
|----------|----------|----------|----------|
| 명사 | 단일어 | 고유 단일어 | 한자단일어 |
| 182~211개 | 142~160개 | 106~115개 | 161~170개 |

| | |
|-----------|-----------|
| 고유 단일어 음절 | 한자 단일어 음절 |
| 100~103개 | 150개 |

- 한국어 비출현 연쇄뿐만 아니라, 제한적으로만 발생하여 회피되는 것으로 가정되는 연쇄도 제약으로 학습됨
 - ✓ 개별 어휘부의 음소 결합 분포가 반영되어 학습됨
 - ✓ 연쇄의 빈도에 따라 가중치가 부여됨

발생 빈도 0인 연쇄에 대한 학습

- 높은 가중치가 부여된 제약으로 학습됨
- ✓ 회피 정도가 강한 제약

| *# $\left[\begin{array}{l} +sonorant \\ +dorsal \end{array} \right]$ (즉, * _w [ŋ]) | | | | | |
|---|------|-----------|-----------|--------------|--------------|
| 명사 | 단일어 | 고유 단일어 | 한자 단일어 | 고유 단일어 음절 | 한자 단일어 음절 |
| 6.48 | 6.48 | 4.29 | 6.38 | 4.05 | 3.22 |

➤ 낮은 가중치의 제약을 다수 학습하여 포착

✓ 예: 고유 단일어

■ 제약 1: *[-sonorant] $\begin{bmatrix} -\text{aspirate} \\ -\text{tense} \end{bmatrix}$ 가중치: 1.15

■ 제약 2: *[-sonorant][-tense] 가중치: 2.54

■ 저해음과 평음의 연쇄(출현 빈도 0) 포착

- [저해음]과 [평음] 및 [격음]의 연쇄를 저지하는 제약(제약 2)에 더해 [저해음]과 [평음]을 금지하는 제약(제약 1)이 학습됨

저빈도 연쇄에 대한 학습

- 상대적으로 낮은 가중치가 부여된 제약으로 학습됨

* $[+labial]$ $\begin{bmatrix} +high \\ +back \\ -round \end{bmatrix}$ (즉, $*[m̩, p̩, p̩', p̩^h]$)

| 명사 | 단일어 | 고유 단일어 | 한자 단일어 | 고유 단일어 음절 | 한자 단일어 음절 |
|------|------|-----------|-----------|--------------|--------------|
| 3.13 | 2.92 | 3.31 | 2.09 | 3.16 | 4.10 |

학습 어휘부별 제약 -단어 학습 문법-

- 모든 어휘부에서 학습되는 제약들이 있는 반면, 일부 제약은 특정 어휘부에서만 학습됨
- 단일어 제약 예(&고유 단일어&한자 단일어 제약)

| 제약 | 의미 | 가중치 | 예외 | 명사 출현 예 |
|-------------------------|--------------|------|--------|----------|
| * [+labial][+dorsal] | [mk, pk'] 회피 | 1.65 | 감각, 잠깐 | 밥공기, 기업가 |

학습 어휘부별 제약 -단어 학습 문법-

➤ 고유 단일어 제약 예

| 제약 | 의미 | 가중치 | 예외 | 한자어 출현 예 |
|---------------------------------|------------|------|----|----------|
| * [-word_boundary][+glottis] | 비어두 [h] 회피 | 3.37 | 사흘 | 방학, 구호 |

➤ 한자 단일어 제약 예

| 제약 | 의미 | 가중치 | 예외 | 고유어 출현 예 |
|--|-------------|------|----|----------|
| * [+sonorant] [-low +labial] [+back -round] | [mɪ, mʌ] 회피 | 3.81 | 금언 | 머리, 구멍 |

학습 어휘부별 제약 -음절 학습 문법-

- 최대 세 음소 결합까지 고려하여 학습
 - ✓ 음절 내 회피 음소, 두 음소 결합 회피 제약이 포함됨
 - 한자어 음절 문법 *[+tense]
 - 고유어 음절 문법 *[-anterior] $\begin{bmatrix} +\text{high} \\ +\text{back} \\ -\text{round} \end{bmatrix}$ 즉, *[cɪ, c^hɪ, c'ɪ]
- ✓ 세 음소 결합 회피 제약 학습
 - 이른바 '우연한 빈칸'으로 일컬어지는 음절이 자질 결합 문법으로 포착될 수 있음

➤ 고유 단일어 음절 제약

| 제약 | 의미 | 가중치 | 예외음절 | 한자 단일어 음절 |
|---|--|------|-----------|-----------|
| *[+ <i>aspirate</i>][+ <i>high</i>][- <i>sonorant</i>] | [p ^h , t ^h , c ^h][i, i, u][p, k] 회피 | 1.82 | 축 (뒤축) | 칙, 특 |

➤ 한자 단일어 음절 제약

| 제약 | 의미 | 가중치 | 예외음절 | 고유 단일어 음절 |
|---|-----------------------------|------|------|-----------|
| *[+ <i>coronal</i>][+ <i>round</i>][+ <i>labial</i>] | [t, s, c][o, u][p, m] 회피 | 1.12 | - | 숨, 줌, 돔 |

단어 적형성 판단 조사

➤ 목적

- ✓ 한국어 화자의 단어 적형성 인식이 학습된 문법들의 예측과 얼마나 일치하는지를 살펴 보고자 함

➤ 조사 대상자

- ✓ 서울/경기 지역 출신 화자 139명
(타 지역에서 4년 이상 거주한 경우 제외)

➤ 조사 방법: 리커트 척도(1-7점) 방법 채택

- ✓ 각 응답자가 컴퓨터 화면에 제시된 단어를 보고, 한국어다움을 판단하도록 함
- ✓ 온라인 설문조사 프로그램 쉐트릭스(Qualtrics) 이용

- 조사단어 (1): 단음절어 819개 [기존단어 포함]
 - ✓ 제약 위배 형태 위주
 - 구성: [자음] - [모음] - [자음], [자음] - [활음] - [모음] 위주
 - 예: 킱, 빙, 줄, 출, 규, 뷰

- 조사단어 (2): 2음절어 600개 [실제 599 단어 분석]
 - ✓ 단음절어에서 확인할 수 없는 [자음] - [자음], [모음] - [모음], [모음] - [격음] 제약 위배를 확인
 - ✓ 기존 단음절어 조합 & 단음절어가 아닌 음절을 조합: 500개
 - ✓ 비출현 음절을 무작위로 조합: 100개
 - ✓ 조사단어(비단어) 구성 예

| 위배 연쇄 | 조사단어 | 1음절의 실제 단어 여부 | 2음절의 실제 단어 여부 |
|-----------------|------|------------------|------------------|
| pt ^h | 밥턱 | YES | YES |
| | 핍퐁 | NO | NO |

조사과정

‘한국어 기준’ 제시: 실제 2음절어 20개

고유 단일어 조건

명사 조건

※ 사용 빈도가 높은 2음절어 제시

단어성 판단 조사(1-7점)

2음절어 조사

단음절어 조사

- ✓ 기준: "닥[닥]" 은 전혀 한국어답게 들리지 않습니다.
이와 같은 단어에 1점을 줄 수 있습니다.
- ✓ 실제단어 또는 새로운 단어 100개를 보여드립니다.
각 단어의 발음을 고려하여, 한국어다움을 판단해 주세요.
1점- 절대 한국어 단어처럼 들릴 수 없다.
4점- 한국어 단어일 수 있지만, 이상하게 들릴 것이다.
7점- 전형적인 한국어 단어다.

조사화면

➤ 기존단어 제시

나는 "**국내[궁내]**"(을/를) 말해.

➤ ‘한국어다움’을 판단; 괄호 안 [발음] 기준

다음 "**단어 [발음]**"의 한국어다움을 판단해 주세요.
(1점-불가능, 7점-전형적)

나는 "**핍칙[핍칙]**"(을/를) 말해.

1

2

3

4

5

6

7

결과 분석

- 개별 어휘부가 응답 점수에 영향을 미치는 영향을 확인하고자 함
- 고유 단일어 제시 조건과 명사 제시 조건의 결과를 통합

[1] 제시 단어에 대한 **응답 점수의 평균**

- 제약을 위배하는 형태(위배형)가 제약을 위배하지 않는 형태(통제형)보다 응답 점수가 더 낮을 것이 예측됨

[2] 제시 단어의 **응답 점수 평균**과 **비문법성 점수**의 상관관계

- 제시 단어가 한국어답지 않을수록, 응답 점수가 낮음
- 제시 단어의 비문법성 점수가 높을수록 한국어답지 않을 것이 예측됨
- 비문법성 점수와 응답 점수는 **‘음의 상관 관계’**를 보일 것으로 예측됨

[3] 혼합 효과 선형 모델

- 어떤 어휘부의 비문법성 점수가 응답 점수에 유의미한 영향을 미치는지를 종합적으로 보고자 함
- 조사단어와 응답자의 임의 효과를 고려한 통계 모델

분석 자료의 예

- 응답 점수
- 조사단어가 개별 어휘부의 문법을 위반하는 정도에 따라 비적형성 점수를 받음

➤ 단음절어

| 조사단어 | 응답자 | 응답 점수 | 명사 | 단일어 | 고유 단일어 | 한자 단일어 | 단일어 | 고유단일어 음절 | 한자단일어 음절 |
|------|-------|-------|-------|-------|-----------|-----------|-------|-------------|-------------|
| 각[각] | sub10 | 1 | 0 | 0 | 0 | 0 | 0 | 6.726 | 2.897 |
| 각[각] | sub56 | 2 | 0 | 0 | 0 | 0 | 0 | 6.726 | 2.897 |
| 븐[븐] | sub28 | 2 | 5.007 | 5.412 | 3.311 | 5.055 | 5.412 | 3.164 | 4.103 |
| 븐[븐] | sub38 | 3 | 5.007 | 5.412 | 3.311 | 5.055 | 5.412 | 3.164 | 4.103 |

➤ 2음절어

| 조사단어 | 1 | 2 | 응답자 | 응답점수 | 고유 단일어 | 한자 단일어 | 단일어 | 명사 | 고유단일어 음절 | 한자단일어 음절 |
|------|---|---|-------|------|-----------|-----------|-------|-------|-------------|-------------|
| 벽출 | 벽 | 출 | sub22 | 1 | 2.544 | 1.615 | 0 | 0 | 3.758 | 6.696 |
| 핍퐁 | 핍 | 퐁 | sub2 | 1 | 2.544 | 3.546 | 5.223 | 3.496 | 7.803 | 19.08 |

✓ 개별 음절 학습 문법이 예측한 2음절어의 비적형성 점수

- 개별 음절 학습 문법이 1음절과 2음절에 부여한 비적형성 점수의 합

| 조사단어 | 1 | 2 | 고유단일어 음절 | 1음절_ 고유단일어 음절 | 2음절_ 고유 단일어 음절 |
|------|---|---|-------------|------------------|-------------------|
| 벽출 | 벽 | 출 | 3.758 | 0 | 3.758 |
| 조사단어 | 1 | 2 | 한자단일어 음절 | 1음절_ 한자단일어 음절 | 2음절_ 한자단일어 음절 |
| 벽출 | 벽 | 출 | 6.696 | 3.844 | 2.852 |

단음절어 조사 결과: 평균

➤ 제약을 위반하는 음절(위배형)이 제약을 위반하지 않는 음절(통제형)보다 낮은 응답 점수를 받음.

✓ 핍 응답 점수: 3.67 * [+labial] [-back] [+labial] 위배
cf. 검 응답 점수: 4.83

| 단어 학습 문법 | 음절 학습 문법 | 평균 | 예 | | | |
|-------------|-------------|------|---|---|---|---|
| 허용 | 허용 | 6.20 | 면 | 겁 | 검 | 슴 |
| 위배 | 허용 | 4.98 | 흙 | 언 | 칭 | 싱 |
| 허용 | 위배 | 4.90 | 득 | 밤 | 핍 | 갈 |
| 위배 | 위배 | 4.11 | 층 | 틀 | 눌 | 몹 |

단음절어 조사 결과: 상관 관계

- 조사단어별 응답 점수 평균과 어휘부별 비문법성 점수의 상관 관계
 - ✓ 약한 음의 상관 관계를 보임
 - ✓ 음절 학습 문법이 단어 학습 문법보다 상관 관계가 다소 높다.

| 단어 학습 문법 | 상관 계수 | 음절 학습 문법 | 상관 계수 |
|----------|-------|--------------|-------|
| 명사 | -0.45 | 고유 단일어 음절 | -0.55 |
| 단일어 | -0.51 | | |
| 고유 단일어 | -0.45 | 한자 단일어 음절 | -0.57 |
| 한자 단일어 | -0.43 | | |

2음절어 조사 결과: 평균

- 단음절어에 비해, 상대적으로 응답 점수가 낮음
- 어떤 제약도 위배하지 않는 형태(통제형)는 다른 위배형보다 높은 응답 점수를 받음

| 단어 문법 | 음절 문법 | 평균 | 예 |
|-----------|-----------|------|--------------|
| 허용 | 허용 | 4.25 | 판개 (4.55) |
| 위배 | 허용 | 3.86 | 풀정 (3.82) |
| 허용 | 위배 | 3.64 | 벵든 (3.6) |
| 위배 | 위배 | 3.05 | 챗든 (2.88) |

2음절어 조사 결과: 상관 관계

- 비단어별 응답 점수 평균과 어휘부별 비문법성 점수의 상관 관계
 - ✓ 약한 음의 상관 관계 확인 cf. 고유 단일어
 - ✓ 단음절어 조사에 비해, 상관 관계가 낮음
 - ✓ 음절 학습 문법이 단어 학습 문법보다 상관 관계가 다소 높음

| 단어 학습 문법 | 상관 계수 | 음절 학습 문법 | 상관 계수 |
|----------|---------------|--------------|-------|
| 명사 | -0.17 | 고유 단일어 음절 | -0.35 |
| 단일어 | -0.24 | | |
| 고유 단일어 | -0.12(p=0.03) | 한자 단일어 음절 | -0.48 |
| 한자 단일어 | -0.24 | | |

혼합효과 선형 모델

➤ 단음절어 조사와 2음절어 조사를 통합

✓ lmer(응답점수 ~

+조사단어의 유형+명사+단일어+고유 단일어+한자 단일어+고유 단일어 음절+한자 단일어 음절+(1|조사단어)+(1|응답자))

➤ 유의미한 효과만으로 통계 모델 구성

| | Estimate | Std. Error | df | t value | Pr(> t) |
|--------------|----------|------------|---------|---------|-----------|
| (Intercept) | 5.76 | 0.08 | 303.51 | 69.90 | <0.001*** |
| 조사단어유형(2음절어) | -1.40 | 0.06 | 1366.98 | -22.06 | <0.001*** |
| 고유 단일어 | -0.11 | 0.02 | 1371.40 | -6.00 | <0.001*** |
| 고유 단일어 음절 | -0.24 | 0.02 | 1381.95 | -14.34 | <0.001*** |
| 한자 단일어 음절 | -0.11 | 0.01 | 1367.36 | -18.37 | <0.001*** |

- 세분화된 어휘부가 적형성 판단에 영향을 끼치는 것으로 보임
 - ✓ 세분화된 어휘부를 대상으로 제약을 탐색할 필요를 시사함

- 개별 어휘부의 상대적 기여도(잠정적)
 - ✓ 고유 단일어 음절 > 고유 단일어 ≍ 한자 단일어 음절

요약

학습 어휘부를 세분화하여, 음소배열제약을 학습

- 단어 단위 문법뿐만 아니라, 음절 단위의 문법을 포함
- 비범주적 인식을 포착할 수 있는 제약을 학습

단어성 판단 조사

- 화자가 학습 문법이 예측하는 비문법성을 인식하는지를 점검
- 개별 어휘부의 영향을 부분적으로 확인

남은 과제

개별 어휘부별 제약 점검

- 어휘부별 동일하거나 유사한 제약이 학습됨
- 제약의 일반화가 필요할 수 있음

제약 선별

- 화자들이 학습된 모든 제약을 인식하는 것으로 보이지 않음

적형성 인식에 영향을 미치는 기타 요인

감사합니다.

