Frequency effects and regularization in Korean nouns

Yoonjung Kang
University of Toronto Scarborough
kang@utsc.utoronto.ca

1. Introduction

(1) Overview

(Kwak 1984, J. Choi 1986, Ko 1989, AKS 1990-1995, H. Kang 1992, Hayes 1998, H. Sohn 2001, Albright 2002, 2005, Y. Kang 2002, 2003, 2005, K-J. Lee 2002, NAKL 2004, E. Kang et al. 2004, S. Park 2006, Davis and Kang 2006, Jun 2007, Jun and Lee 2007 among others)

- In many Western dialects of Korean, noun-final coronal obstruents are optionally realized as [**s**]. For /t^h/-final nouns, there is an additional variant with final [**c^h**].

/cə**c**-il/	cə**c**-il	~	cə**s**-il			‘milk, acc.’
/tə**c^h**-il/	tə**c^h**-il	~	tə**s**-il			‘trap, acc.’
/p^{h}a**t^h**-il/	p^{h}a**t^h**-il	~	p^{h}a**s**-il	~	p^{h}a**c^h**-il	‘red bean, acc.’

- These novel variants are not equally available in all suffixal contexts, often creating a "mixed paradigm".

/pat^h/ ‘field’	Acc.	Dir.	Loc.
(H. Kang 1993)	**pac^h-il**	pat^h-ilo	pat^h-e

(2) Proposal

I. Reflection of the frequency distribution in the lexicon (Jun 2007)
II. Phonological reanalysis/overgeneralization
III. Morphological generalization

2. Background

(3) Korean phoneme inventory (cf. H. Kim 1999)

p, p^h, p’	t, t^h, t’	k, k^h, k’	
	c, c^h, c’		
	s, s’		h
m	n	j	
	l		

i	ɨ	u
e	ə	o
(æ)	ɑ	

(4) Coda neutralization: /t^h, t, c, c^h, s, s'/→ [t]/_]$_{coda}$

STEM-DECL.	STEM-and	
[kɑt^h-ɑ]	[kɑt-k'o]	'same'
[pɑt-ɑ]	[pɑt-k'o]	'to receive'
[c'oc^h-ɑ]	[c'ot-k'o]	'to chase'
[cɑc-ɑ]	[cɑt-k'o]	'frequent'
[is'-ə]	[it-k'o]	'to exist'
[pəs-ə]	[pət-k'o]	'to take off'

(5) Affrication: /t, t^h/ → [c, c^h]/__]$_{stem}$(h){i, j} (Derived environments only)

/i/:	/t^hək-pɑt-i/	[t^həkp'ɑci]		'bib'
	/kɑt^h-i/	[kɑc^hi]		'together'
/ɨ/:	/pɑt-ɨni/	[pɑtɨni]	*[pɑcɨni]	'receive-therefore'
	/kɑt^h-ɨni/	[kɑt^hɨni]	*[kɑc^hɨni]	'same-therefore'
/ɑ/:	/pɑt-ɑ/	[pɑtɑ]	*[pɑcɑ]	'receive-IMPERATIVE'
	/kɑt^h-ɑ/	[kɑt^hɑ]	*[kɑc^hɑ]	'same-IMPERATIVE'
/ə/:	/kət-ə/	[kətə]	*[kəcə]	'gather-IMPERATIVE'
	/put^h-ə/	[put^hə]	*[puc^hə]	'stick-IMPERATIVE'

3. Neutralization and variation

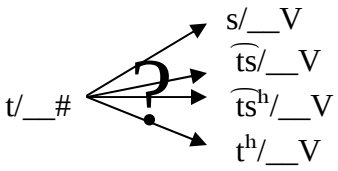
(6) [s]-variants: Analogy to a dominant pattern (Ko 1989, Hayes 1998, Albright 2002, 2005)

- The unsuffixed form is the most "informative" form of the Korean noun paradigm, for the language in general, from which other forms in the paradigm can be projected (Albright 2005).

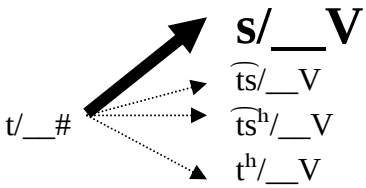
- Neutralization of underlying contrast in noun paradigm

	Unsuffixed	NOM. (-i)	ACC. (-ɨl)	DIR. (-ɨlo)	LOC. (-e)
/-s/	**-t**	-s	-s	-s	-s
/-c/	**-t**	-c	-c	-c	-c
/-c^h/	**-t**	**-c^h**	-c^h	-c^h	-c^h
/-t^h/	**-t**	**-c^h**	-t^h	-t^h	-t^h
	:Coda Neutralization	:Affrication			

- In the unsuffixed form, all coronal obstruents neutralize. Therefore, the learners are presented with the unsuffixed form and have to "guess" what the underlying consonant should be.



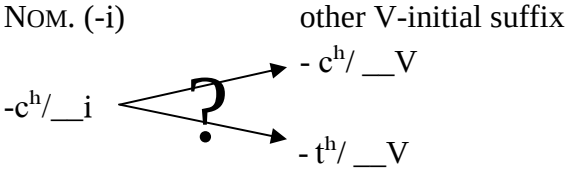
- They opt for the most common /s/-final nouns, which gives them the best chance at being correct.



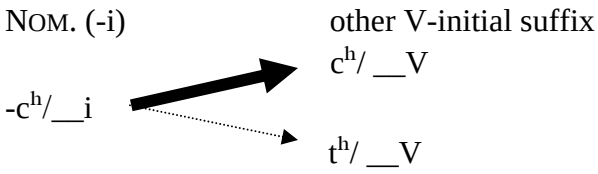
(7) [t^h] ~ [c^h] variation: projecting from the nominative (Kang 2005)

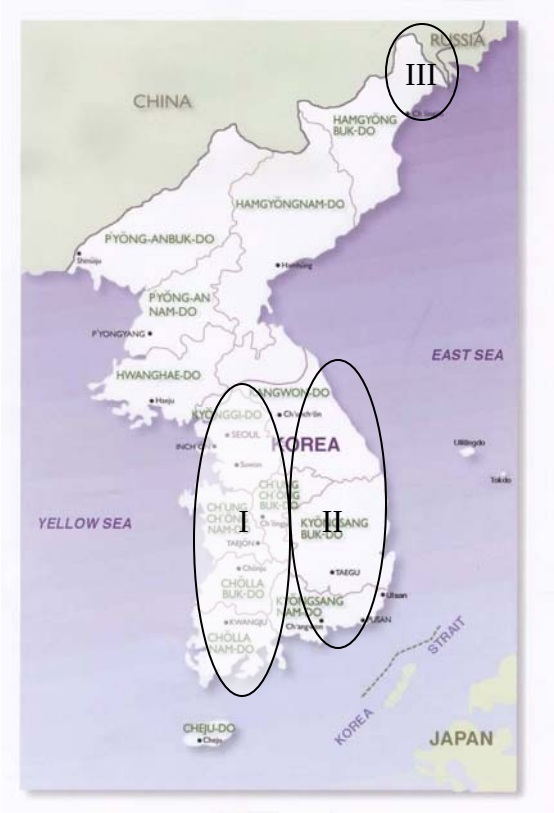
- The nominative form is the second most informative form of the Korean noun paradigm, for the language in general (Albright 2005).

- Given the [c^h]-final form in the nominative, learners cannot be sure which consonant should appear before other vowel-initial suffixes.



I. Western dialects (Seoul, Kyenggi, Chunchung, Cenla): the change is mainly in the direction of original /t^h/ → [c^h] and the change of original /c^h/ → [t^h] is marginal.





II. Eastern dialects (Kangwon, Kyengsang): the change is mainly in the direction of original /c^h/ → [t^h] and the change of original /t^h/ → [c^h] is marginal.

(M. Choi 1980, AKS 1990-1995, Hong 2003, B. Kim 2005)

NOM. (-i) other V-initial suffix

-c^h/__i → c^h/ __V

-c^h/__i → t^h/ __V

III. Hampuk dialect: The dialect does not have affrication rule and therefore, there is no neutralization of /t^h/ and /c^h/ nouns in the nominative form. [t^h] ~ [c^h] variation is not found in either direction (B. Kim 2005).

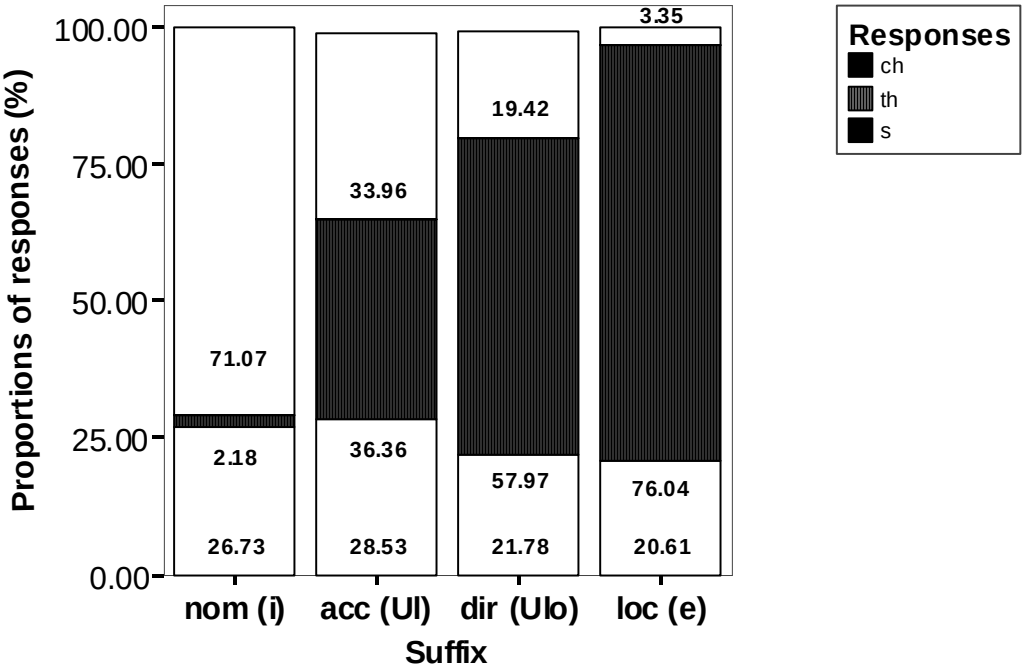
.

4. Distribution of [c^h] variants

(8) The National Academy of Korean Language (2004)

- Hye-Won Choi, Min-Kyeng Suh, Yen-Sin Hwang, Mi-Yeng Kwen
- 1174 Seoul-Incheon-Keynggi speakers
- Multiple choice questionnaire
- 14 /t^h/-final nouns
    - o *pyət^h* 'sunlight', *mit^h* 'bottom', *pat^h* 'field', *k'it^h* 'end', *p^{h}at^h* 'red bean', *sot^h* 'pot', *kət^h* 'outside' *kyət^h* 'side' *nat^h* 'a unit', *twik'jət^h* 'backyard', *məlimat^h* 'bedside', *mut^h* 'land', *pak'at^h* 'outside', and *sut^h* 'thickness (of hair)'

(9) Proportion of [t^h], [c^h] and [s] responses for /t^h/-final nouns in each suffix context (Based on NAKL 2004)



- The proportion of [s] responses is roughly constant across suffixes.
- The ratio between [t^h] and [c^h] responses differs significantly by the suffix.

5. Lexical diffusion and frequency of use

(10) Certain sound changes affect low- and high-frequency words differently
(Schuchardt 1885, Fidelholtz 1975, Hooper 1976, , Phillips 1980, 1983, 1984, 2001, Rhodes 1996, Bybee 1985, 1995, 2001, Labov 1994, Bybee and Hopper 2001, Jurafsky et al 2001, Anttila 2006 among many others)

I. <u>Phonetically motivated changes</u>

- Phonetically motivated changes (typically ones arising from lenition) affect high-frequency words first.
- Phonetically motivated changes progress with each use of the word. Therefore, the more frequent a word is put to use, the more advanced the change is.
  ex. Schwa deletion in English : *ev**<u>e</u>**ry* vs. *mamm**<u>a</u>**ry*

II. <u>Analogically motivated changes</u>

- Analogically motivated changes affect low-frequency words first.
- High-frequency words form strong mental representations and resist change motivated by analogy to other forms.
  ex. Regularization of English irregular past: *weeped (< wept)* vs. **keeped* (<*kept*)

(11) The frequency of use in coronal obstruent nouns

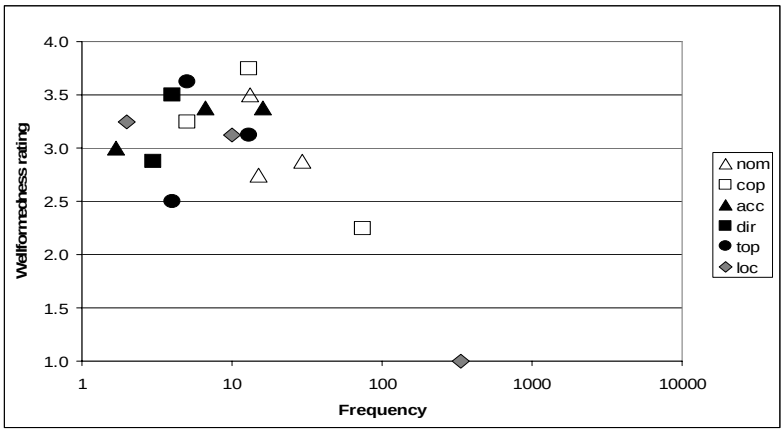
- Nouns with meaning of location or time—which include most /t^{h}/-final nouns—are used in the locative form frequently and the locative form of these nouns tends to resist the changes in noun-final coronals.

‘place’	acc.	loc.	
15C:	*ko**t**-ɔl*	*ko**t**-ɔy*	(C. Kwak 1984)
19C Cenla Dialect:	*ko**c**-ɪl~ko**s**-ɪl*	*ko**t**-e*	(J. Choi 1986)
20C:	*ko**s**-ɪl*	*ko**s**-e*	

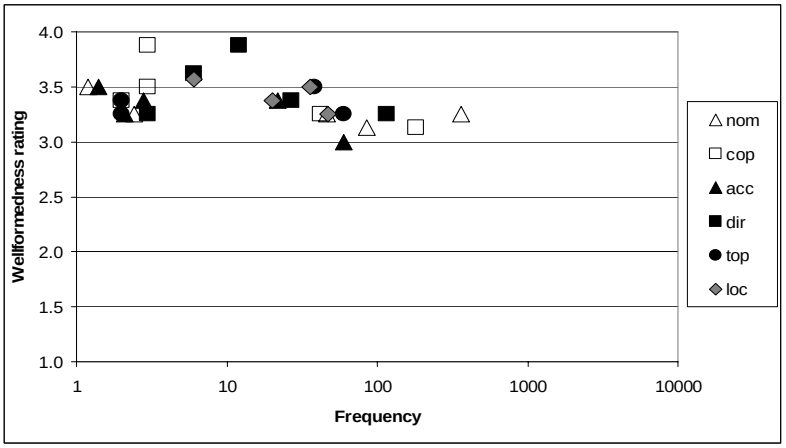
(B. Lee 1975, T. Choi 1977, J. Choi 1986, K. Lee 1986, K. Ko 1989, H. Kang 1993, Y. Kang 2003, 2005)

(12) Kang (2003)

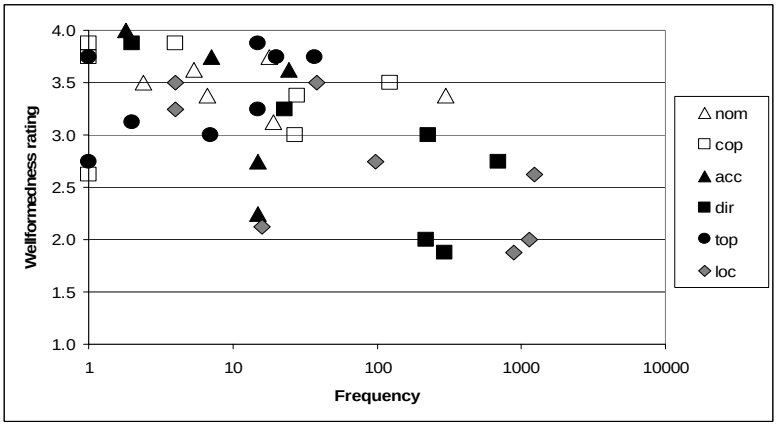
- 8 Seoul speakers; well-formedness rating; Written questionnaire format
- Rating: between 1 (bad) and 4 (good).
- Ratings for **[s] variant** negatively correlate with the frequency of use.[1]
  - /c/-final nouns (3 nouns)



  - /c^h/-final nouns (5 nouns)

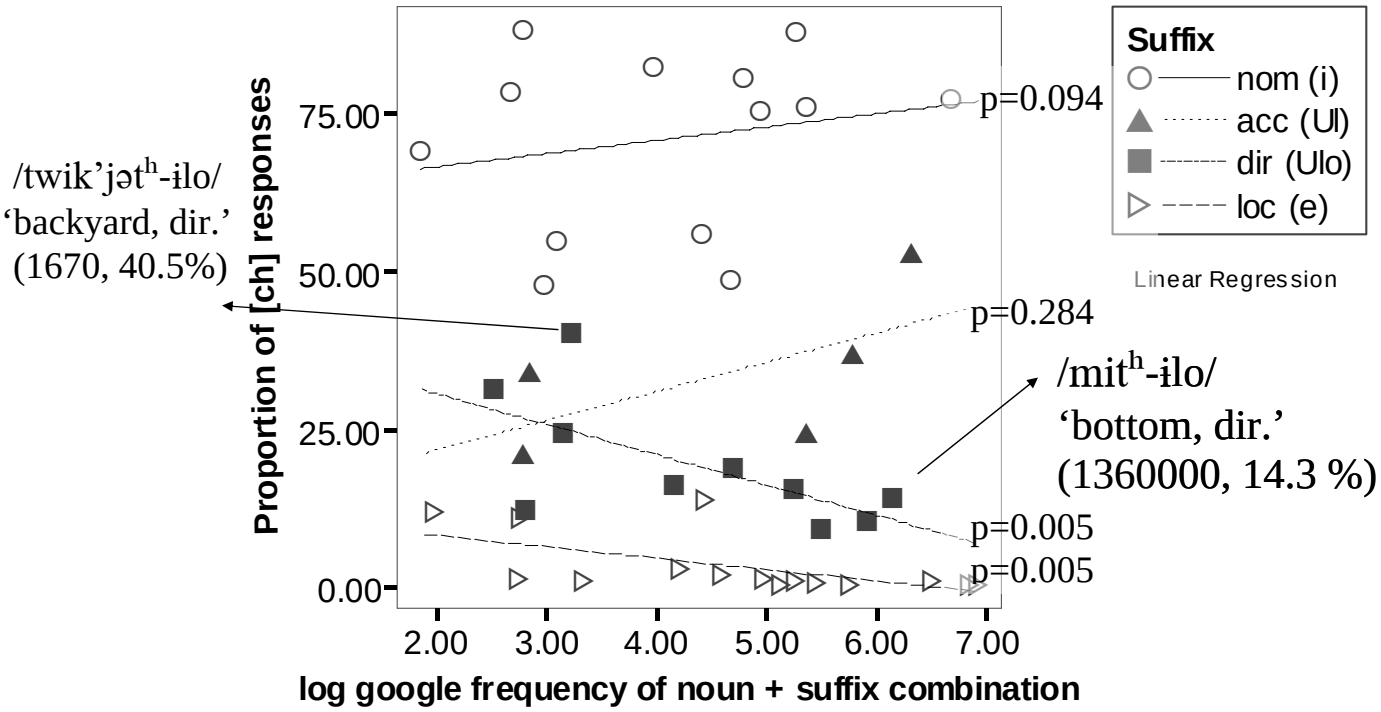


  - /t^h/-final nouns (8 nouns)



[1] Frequency counts from KAIST Concordance program (KCPMSTAT) containing 13.6 million words (http://morph.kaist.ac.kr/kcp/).

(13) The frequency of use and the proportion of [c^h] responses for /t^h/-final nouns (Based on NAKL 2004)



- Frequency counts: Google search (June 2007)
- The frequency of use effect does not explain away the suffix effect: words of similar frequency of use show different rates of [c^h] responses depending on the suffix it contains.

6. Frequency distribution in the lexicon

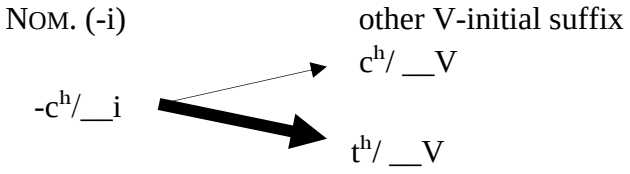
(14) Jun (2007)

- Speakers have a fairly detailed knowledge of statistical distribution of linguistic categories in the lexicon
  (Zuraw 2000, Frisch et al. 2001, Albright 2002, Ernestus and Baayen 2003, Pierrehumbert 2003, Hayes and Londe 2006 among many others)
- Innovative [c^h] variants are more likely to spread to those contexts where /c^h/-final nouns are already abundantly attested.

(15) Google token frequency and NAKL data[2]

	Nom (i)	Acc. (ɨl)	Dir. (ɨlo)	Loc. (e)
Google token frequency Proportion of [c^h] out of [c^h] and [t^h] combined in	**100%** (by rule)	**70.9%**	**22.6%**	**6.2%**
Hits for /c^h/-final nouns		9,263,510	2,059,463	1,313,608
Hits for /t^h/-final nouns		3,810,939	7,070,496	19,772,809
NAKL (2004) Proportion of [c^h]-final responses out of [c^h] and [t^h] final responses combined	**96.8%**	**49.1%**	**27.2%**	**5.5%**

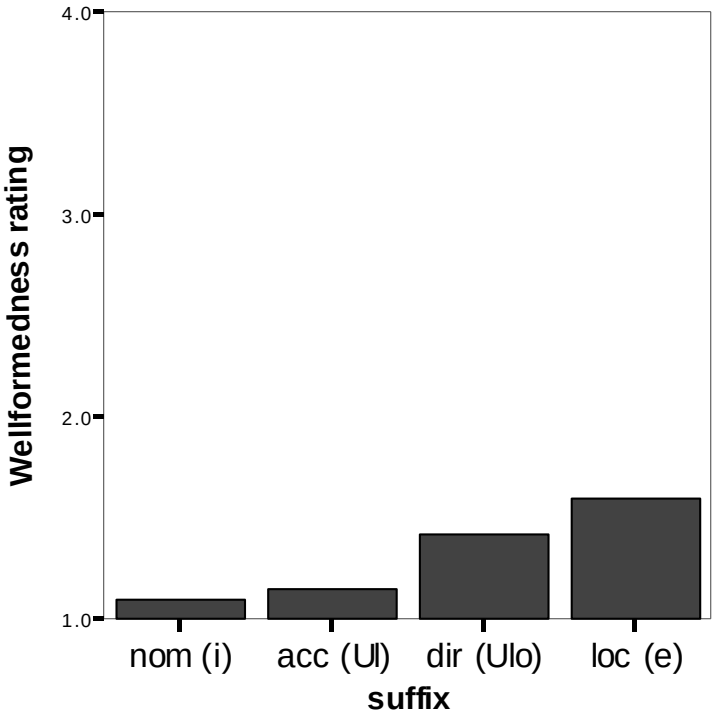
(16) Change in the other direction also honours the hierarchy of relative frequency distribution of [c^h] vs. [t^h] in different suffix context. (Kang 2005)



- Innovative [t^h] variants are more likely to spread to the suffix context where /t^h/-final nouns are already abundantly attested.

[2] Jun (2007) used type frequency of nouns in different suffix contexts in 1.5 million word Sejong corpus. However, the inflectional suffixes productively combine with nouns and given a large enough corpus (such as Google), it is not unlikely that the type frequency of [c^h] vs. [t^h] nouns come out similar across different suffix contexts. For this reason, I am using token frequency counts, rather than type frequency counts here. In this table, the relative frequency counts are based on a Google search (June 2007) of all monomorphemic /t^h/ and /c^h/-final nouns attested in the Sejong corpus: 15 monomorphemic /t^h/-final nouns and 11 monomorphemic /c^h/-final nouns (cf. NAKL 2003 survey of word frequency).

- Kang (2003, 2005) : Ratings for /c^h/ nouns pronounced as [t^h]-final.(Average of 8 speakers)



Google token Frequency Proportion of [t^h] Out of [t^h] and [c^h]	**0 %** (by rule)	**29.1 %**	**77.4 %**	**93.8 %**

- /c^h/ nouns: Kyenggi dialects (AKS 1995)

	Nom./i/	Acc. /ɨl/ (~/əl/)	Loc. /e/	(N=19)
/suc^h/ 'charcoal'	c^h	**c^h**	**t^h**	5
	c^h		t^h	1
	c^h	c^h	c^h	3
	s	s	s	6
	c^h	s	s	1
	c^h	s	c^h	1
	s	s	t^h	1

/k'oc^h/ 'flower'	(c^h)	**c^h**	**t^h**	4
	c^h	c^h	c^h	5
	c^h	c^h	c^h~s	1
	(s)	s	s	8
		s	t^h	1

- /c^h/-final nouns: Southern Kyengsang dialect (AKS 1993)

	Nom./i/	Acc. /əl/	Loc. /e/	(N=15)
/suc^h/ 'charcoal'	c^h	t^h	t^h	17
	c^h	**c^h**	**t^h**	1
	c^h	t^h	c^h	1
/k'oc^h/ 'flower'	c^h	t^h	t^h	15
	c^h	**c^h**	**t^h**	5

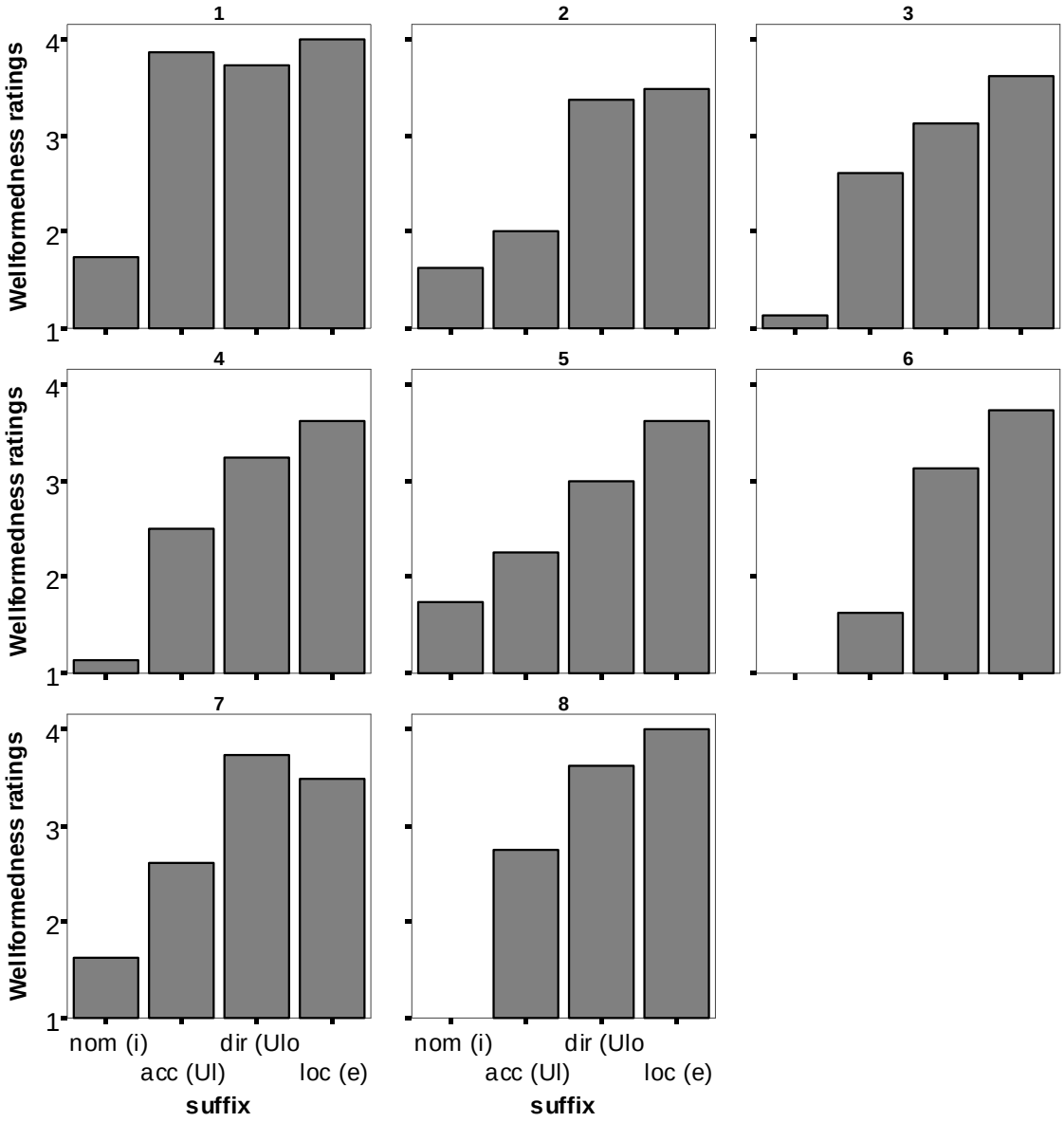
- /c^h/-final nouns: Kangwon dialect (AKS 1990)

	Nom./i/	Acc. /ɨl/	Loc. /e/	(N=15)
/suc^h/ 'charcoal'	c^h	t^h	t^h	7
	c^h	c^h	c^h	4
	c^h	**c^h**	**t^h**	1
	c^h	t^h	c^h	1
/k'oc^h/ 'flower'	c^h	t^h	t^h	8
	c^h	**c^h**	**t^h**	1
	c^h	c^h	c^h	1
	(c^h)	c^h		2
			t^h	1
	c^h~s	t^h		1
	s	s	s	1
/təc^h/ 'trap'	c^h	c^h	c^h	5
	c^h		c^h	1
	c^h	**c^h**	**t^h**	1
	c^h			1
	c^h		t^h	1
	c^h	t^h	t^h	5
	s	s	s	1

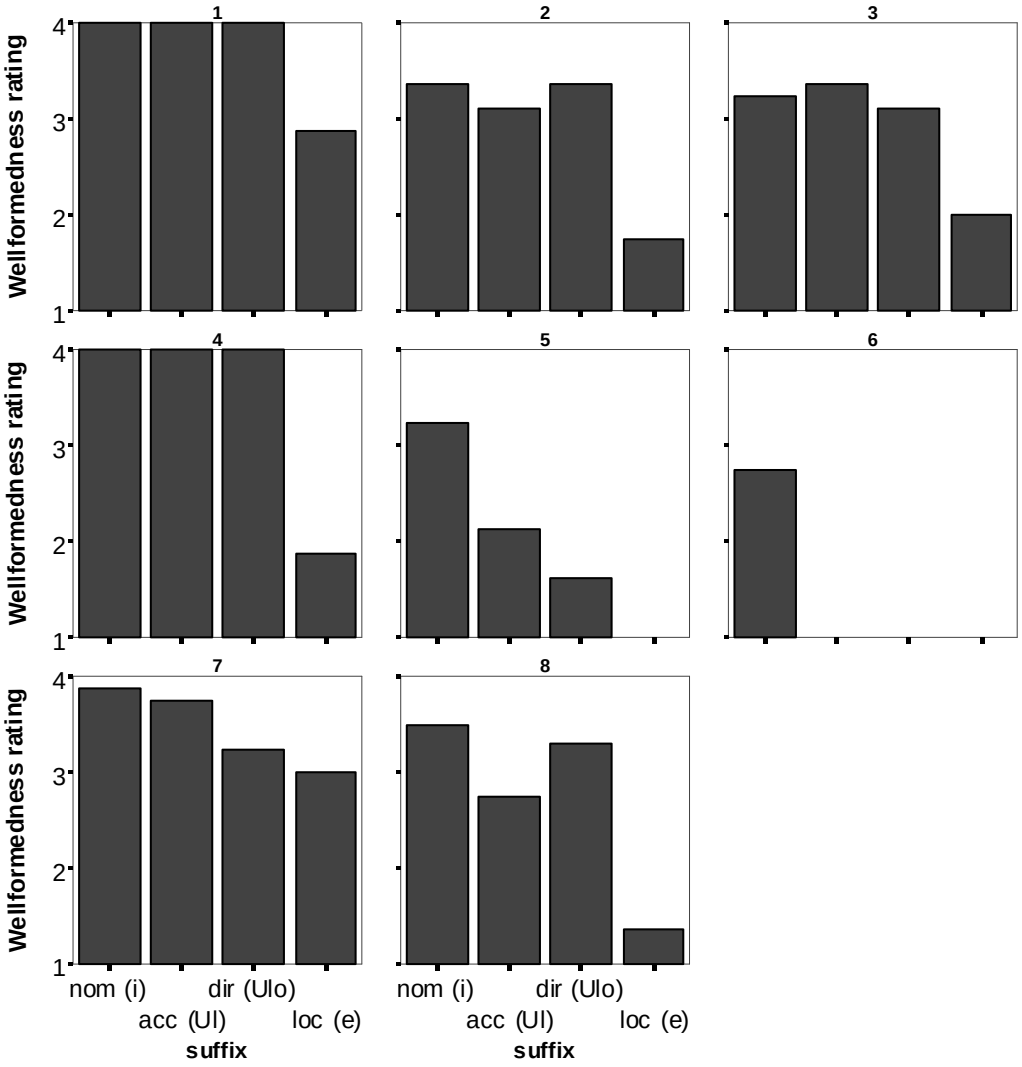
7. Wellformedness rating

(17) Kang (2003, 2005)

- Wellformedness rating for [t^h] pronundication of /t^h/-final nouns
- Mirrors the frequency distribution of [t^h].



- Wellformedness rating for [c^h] pronundication of /t^h/-final nouns
- Some speakers seem to have reanalyzed the suffixal asymmetry as one **conditioned by the vowel height.**



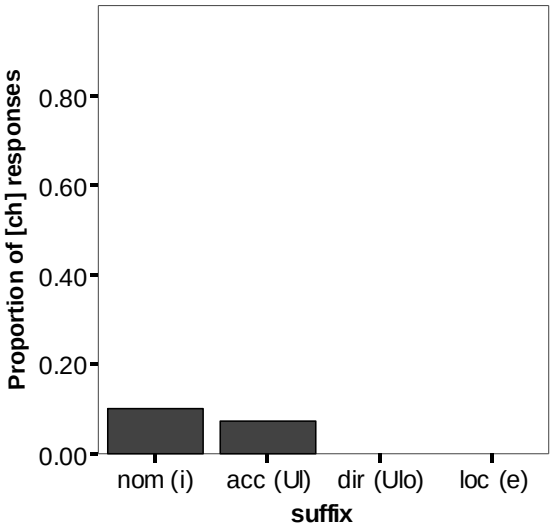
- Speaker 6 does not accept innovative [c^h] variants at all.
- Speakers 5 and 7 show the suffix asymmetry in line with the lexical frequency.
- Four speakers (1, 2, 3 and 4) rate [c^h] variants in all high vowel contexts (Nom, Acc, Dir) comparably good.
- These speakers seem to have reanalyzed the distribution of [c^h] as conditioned by the vowel height and extended the context of affrication rule to all high vowel contexts.

7. Morphonological generalization

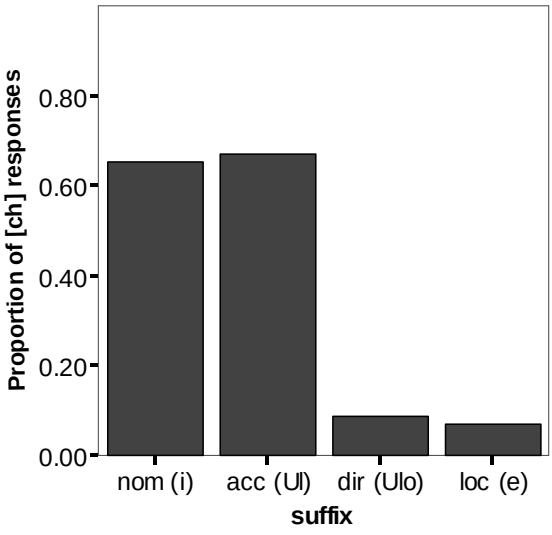
(18) Acc.-Dir. Split in Southern Cenla dialects

- H. Kang (1992): Kwangcwu and Kwangyang dialects
- 32 speakers each

- Kwangcwu



- Kwangyang



- Dir. (ɨlo) patterns with Loc. (e) and shows a very low rate of [c^h] responses, while Acc (ɨl) shows a rate of [c^h] response comparable to that of Nom. (i).

(19) Why no phonological regularization in these dialects?

- Variation in suffixal vowels in many Western dialects (including Cenla) (J. J. Choi 2001, K. Lee 2000, Han and Kim 200x)
  - The locative suffix (-e) is variably realized as /i/ (< ɨj 19th Century).
  - But, ɨj > i change occurred after the affrication (t, t^h > c, c^h/__ i) ceases to be productive and [t^h] fails to affricate in the locative.
    - [mit^h-i] 'bottom, loc.'
    - [pat^h-ida] 'field, loc.'
  - Affrication does not reliably apply before any vowel context. And the learners do not seek a phonological generalization for the distribution of [c^h]. Rather, they resort to the morphological contexts for generalization.

(20) Implications

- Sound changes can be sensitive to the frequency distribution of the lexicon.
- Faced with a complex pattern of variation, learners/speakers seem to seek a simpler generalization (phonological or morphological).
- Phonology remains manageable.

Selected References

한국 정신 문화 연구원 (The Academy of Korean Studies). 1990. 한국방언자료집[Korean dialectal data] 2: 강원도편

한국 정신 문화 연구원 (The Academy of Korean Studies). 1990. 한국방언자료집[Korean dialectal data] 4: 충청남도편

한국 정신 문화 연구원 (The Academy of Korean Studies). 1991. 한국방언자료집[Korean dialectal data] 6: 전라남도편

한국 정신 문화 연구원 (The Academy of Korean Studies). 1993. 한국방언자료집[Korean dialectal data] 8: 경상남도편

한국 정신 문화 연구원 (The Academy of Korean Studies). 1995. 한국방언자료집[Korean dialectal data] 1: 경기도편

Albright, A. 2002. Albright, Adam. 2002. *The identification of bases in morphological paradigms*. UCLA PhD Dissertation.

Albright, A. 2005. Explaining universal tendencies and language particulars in analogical change. Manuscript, MIT.

Antilla, A. 2006. Variation and opacity. *Natural Language and Linguistic Theory.* 24(4), 893-944

Bybee, Joan L. 1985. *Morphology: a study of the relation between meaning and form.* Philadelphia: Benjamin.

Bybee, Joan and Paul Hopper. eds. 2001. *Frequency and the emergence of linguistic structure.* Philadelphia: Benjamin.

Bybee, Joan . 2001. *Phonology and language use.* Cambridge University Press.

최 전승 (Choi, J.) 1986. 19 세기 후기 전라방언의 음운현상과 그 역사성. [Phonology of late 18th century Cenla dialect]. 전북대학교 박사학위 논문.

최 전승 (Choi, J.) 2001. 19 세기 후기 전라방언의 처소격 조사 부류의 특질과 변화의 방향 [Characteristics and changes in locative suffix in late 19th century Cenla dialect]. 우리말글 20.

최 명옥 (Choi, M.) 1980. 경북 동해안방언 연구: 영덕군 영해면을 중심으로 [A study of Northern Kyengsang coastal dialects: with a special focus on Yengtekkwun Yenghaemyen.] 영남대출판부.
Frisch, Stefan A., Nathan R. Large, Bushra Zawaydeh, & David B. Pisoni. 2001. Emergent phonotactic generalizations in English and Arabic. *Typological studies in Language 45.* John Benjamins: Amsterdam. 159-179.
한영목, 김래영. (Han and Kim) 200x. 충남방언 조사의 형태변이와 통사 의미의 몇 양상 [On Inflectional suffixes in Chungnam dialects]. 인문학 연구 30. 2.
Hayes, Bruce. 1998. On the Richness of Paradigms, and the Insufficiency of Underlying Representations in Accounting for them. Handout, Stanford.
Hong, Mi-Ju. 2003. A sociolinguisc anslysis on the use of substantive stem-final consonants (c^h) and (t^h). *The Sociolinguistic Journal of Korea 11.1.*
Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. ed., *Current progress in historical linguistics,* North Holland: Amsterdam, 96-105.
Jun, Jong-Ho. 2007. Variation in Korean stem-final obstruents. Ms, Seoul National University.
Jun, Jong-ho and Jee-Hyun Lee. 2007. Multiple stem-final variants in Korean native nouns and loanwords. Enehag.
Kang, Eunji, Ho-Youn Lee, Juwon Kim. 2004. The phonetic realization of syllable codas in Korean. *Malsori 49.*
강 희숙 (Kang, H.) 1992. 국어 마찰음화에 대한 연구—전남 방언을 중심으로 [On fricativization in Korean—with a focus on Southern Cenla dialect]. 인문과학연구 92-1.
강 희숙 (Kang, H.) 1993. 음운변이 및 변화에 관한 사회언어학적 연구: 전남 장흥 방언을 중심으로 [A sociolinguistic study on sound changes and variations: with a focus on Southern Cenla, Canghung dialect]. 전남대학교 박사학위 논문.
Kang, Yoonjung. 2003. Sound changes affecting noun-final coronal obstruents in Korean. In McClure, W. ed. *Japanese/Korean Linguistics 12.*
Kang, Yoonjung. 2005. The Emergence of the Unmarked in an analogical change. Handout, Seoul Forum.
김 봉국 (Kim, B.) 2005. 체언 어간말 중자음의 편화 양상: 동해안 방언 및 함북 육진 방언을 중심으로 [Changes in noun-final coronal sounds: East coast and Hampuk Yukjin dialects]. 국어학 45.
고 광모. (Ko, K.) 1989. 체언 끝의 변화 ㄷ>ㅅ 에 대한 새로운 해석 [A new analysis on noun-final t > s change]. 언어학. 11 호 3-21.
곽 충구. (Kwak, C.) 1984. 체언 어간말 설단자음의 마찰음화에 대하여 [On fricativization of noun-final coronal sounds]. 국어국문학 91.
국립국어연구원 (The National Academy of Korean Language). 2004. 표준 발음 실태 조사 [A survey of standard pronunciation] 3. Downloadable from http://www.korean.go.kr/
이 길재. (Lee, K-J.) 2002. 언어변이와 사회계층에 대한 일고—나주지역어의 마찰음화를 중심으로 [A study on language variation and social classes—fricativization in Naju dialect]. 한국언어문학 55.
이 기갑. (Lee, K-K.) 2000. 국어 방언의 조사체계 [Inflectional suffixes in Korean dialects]. 언어학 [Enehag] 27.
Park, Sunwoo. 2006. *Paradigm Uniformity Effects in Korean Phonology*. PhD. Dissertation, Korea University, Seoul, Korea.
Phillips, Betty. 1983. Lexical diffusion and function words. *Linguistics 21,* 487-99.
Phillips, Betty. 1984. Word frequency and the actuation of sound change. *Language 60,* 320-42.
Phillips, Betty. 2001. Lexical diffusion, lexical frequency, and lexical analysis. In Bybee, Joan and Paul Hopper eds., *Frequency and the emergence of linguistic structure.*
Schuchardt, Hugo. 1885. On sound laws: against the neogrammarians. English translation by Vennemann, T. and T. H. Wilbur [1972].