Variable adaptations of English word-final stops by Koreans in Stochastic OT

Na-Young Ryu & Sung-Hoon Hong

Hankuk University of Foreign Studies
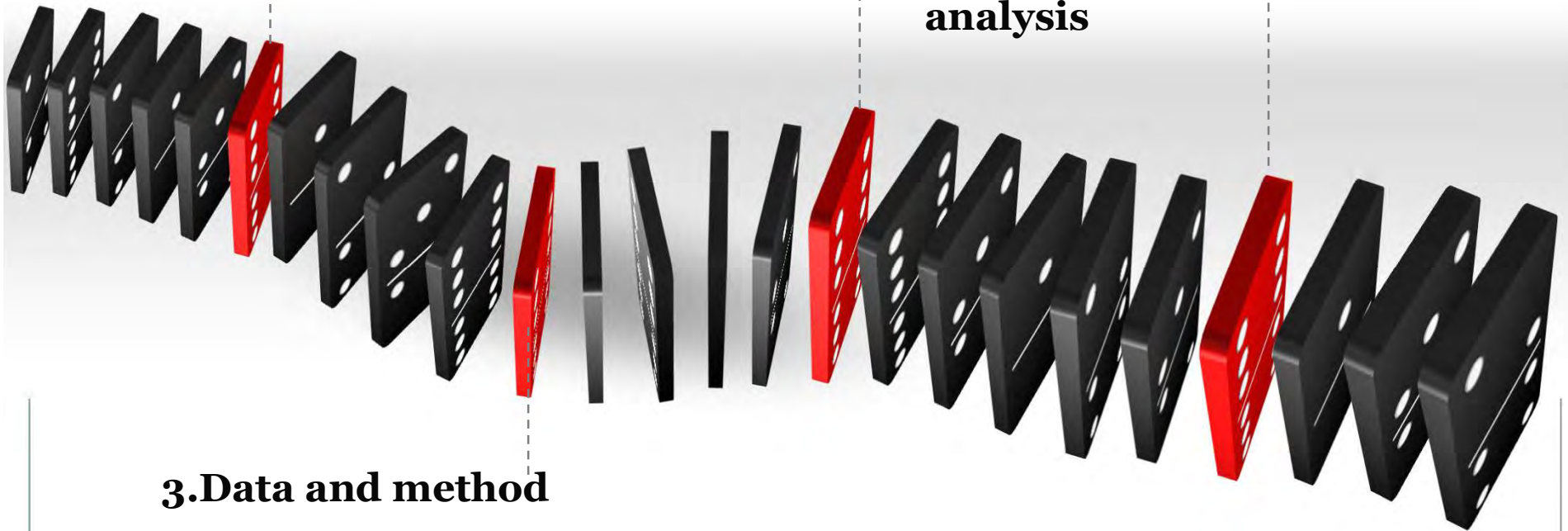
Table of contents

1.Introduction

- Investigate variable adaptations of English word-final stops by Koreans.

- Follow both a linguistic and sociolinguistic approach based on a survey of 30 tokens.

- Predict the likelihood of coda production and vowel epenthesis in the adaptations based on the factor weight.

- Adopt a stochastic version of Optimality Theory for the analysis of the variability of word-final stops in English loanwords.

Variable adaptations of English loanwords

[ɨ] Vowel epenthesis	[Ø] Coda production	[Ø]~[ɨ] Variation
'light' → [la.i.tʰɨ]	'group' → [kɨ.lup]	'soup' → [sup] ~ [su.pʰɨ]
'red' → [lɛ.dɨ]	'bag' → [pæk]	'hip' → [hip] ~ [hi.pʰɨ]
'leage' → [li.gɨ]	'comic' →[kʰo.mik]	'cassette' → [kʰa.sɛt] ~ [kʰa.sɛ.tʰɨ]

2. Survey

- Purpose: To show how Koreans adopt English loanwords ending in word-final stops according to linguistic and extra-linguistic factors.

- Data: 30 English word-final stops in Korean (NIKL 2005)

Vowel epenthesis preferred	5	(ex: league, site, merit)

Coda Production preferred	5	(ex: target, internet, group)

Variation preferred	20	(ex: jeep, soup, cake, tape)

- Participants: 505 (online: 130, offline: 375)
- Online survey: http://maincc.hufs.ac.kr/~hongsh
- Period: 2010.04.05~05.10

3. Data and Method

Data: The results of the survey

Method: GOLDVARB X program (Sankoff et al 2005)

(http://individual.utoronto.ca/tagliamonte/goldvarb.htm)

- GOLDVARB X is a **tool of historical linguistics and sociolinguistics** to explicate patterns of variation between alternative forms in language use.

- This program can be used with **linguistic and extra-linguistic factors** that results in probabilities of rule application.

- A variable rule analysis computes **a multivariate statistical model,** on the basis of observed token counts.

- Each determining factor is assigned **a numerical factor weight** that describes how it influences the probabilities of choice of either form.

# 4. Linguistic and extra-linguistics factors

1.Known linguistic factors

(1) Tenseness of the pre-final vowel
  (tense vs. lax)
(2) Voicing of the final stop
  (voiced vs. voiceless)
(3) Place of articulation
  (coronal vs. labial & dorsal)
(4) The number of syllables
  (monosyllabic vs. polysyllabic)

(Hirano 1994, O.Kang 1996, P.Lee 1998, E.Jun 2002, Y.Kang 2003)

2.Possible extra-linguistic factors

(1) Level of English proficiency
  (beginner vs. intermediate vs. advanced)
(2) Speech style
  (formal vs. informal)
(3) Age

(cf.cardoso 2007)

Linguistic factors

Extra-linguistic factors
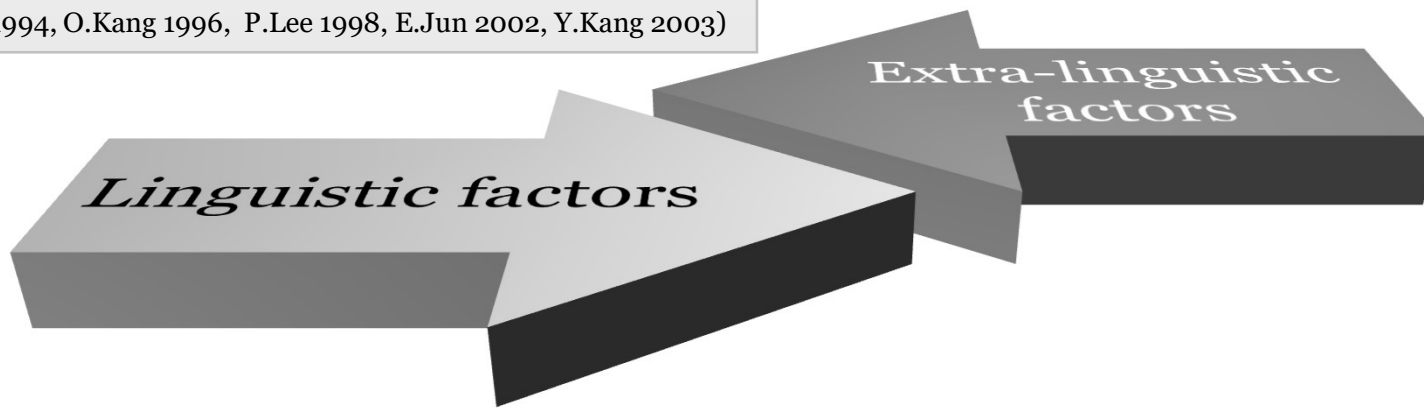
Factor groups for GOLDVARB X analysis

Dependent Variables	a. Coda production	b. Vowel epenthesis

Linguistic factor groups

(1) Tenseness of the pre-final vowel	a. tense b. lax
(2) Place of articulation	a. coronal b. labial c. dorsal
(3) The number of syllables	a. monosyllable b. Polysyllable
(4) Voicing of the final stop	a. voiced b. voiceless

Extra-linguistic factor groups

(5) Age	a. under 20 b.21~30 c. 31~40 d. 41~50 e. over 50
(6) The length of residence in English speaking countries	a. none b. less 1~5years c. more than 5 years
(7) TOEIC score	a. none b. less 500 c.501~700 d.701~900 e.901~990
(8) English proficiency	a. beginner b. intermediate c. upper-intermediate d. advanced
(9) Frequency	a. low frequency b. high frequency

Results of step-up & down regression analaysis

Significant Factor groups

1. Tenseness of the pre-final vowel
2. Place of articulation
3. The number of syllable
4. Voicing of the final stop
5. Length of the residence in English speaking countries
   (All factors are statistically significant p<.001)

Non-significant Factor groups

1. Age [Log likelihood=-8676,698 p=0.045]
2. TOEIC score [Log likelihood =-9941.708 p=0.447]
3. English proficiency [Log likelihood=-699.196 p=0.004]
4. Frequency [Log likelihood=-9167.673 p=0.469]

•Log likelihood: measure of the goodness of fit of an analysis; figures closer to zero represent better models than those further removed from zero.

5. Quantitative results

Final GOLDVARB X probabilistic results

Factor groups	Likelihood of coda production		
Tenseness of pre-final vowel	Tense (0.38)	**Lax (0.59)**	
Place of articulation	Coronal (0.41)	**Labial (0.67)**	**Dorsal (0.72)**
The number of syllables	Monosyllabic (0.40)	**Polysyllabic (0.68)**	
Voicing of the final stop	Voiced (0.08)	**Voiceless (0.54)**	
The length of residence in English speaking countries	Level1	Level2	Level3

Chi-square/cell:25.0139

5.Quantitative results

• **Significant factors affecting of coda production and vowel epenthesis**

Coda production	Vowel epenthesis
Live abroad for over 5 years (0.71)	Voiced (0.92)
Dorsal & Labial (0.72 & 0.67)	Tense (0.62)
Polysyllabic (0.68)	Monosyllabic (0.60)
Lax (0.59)	Coronal (0.59)
Voiceless (0.54)	Only live in Korea (0.53)
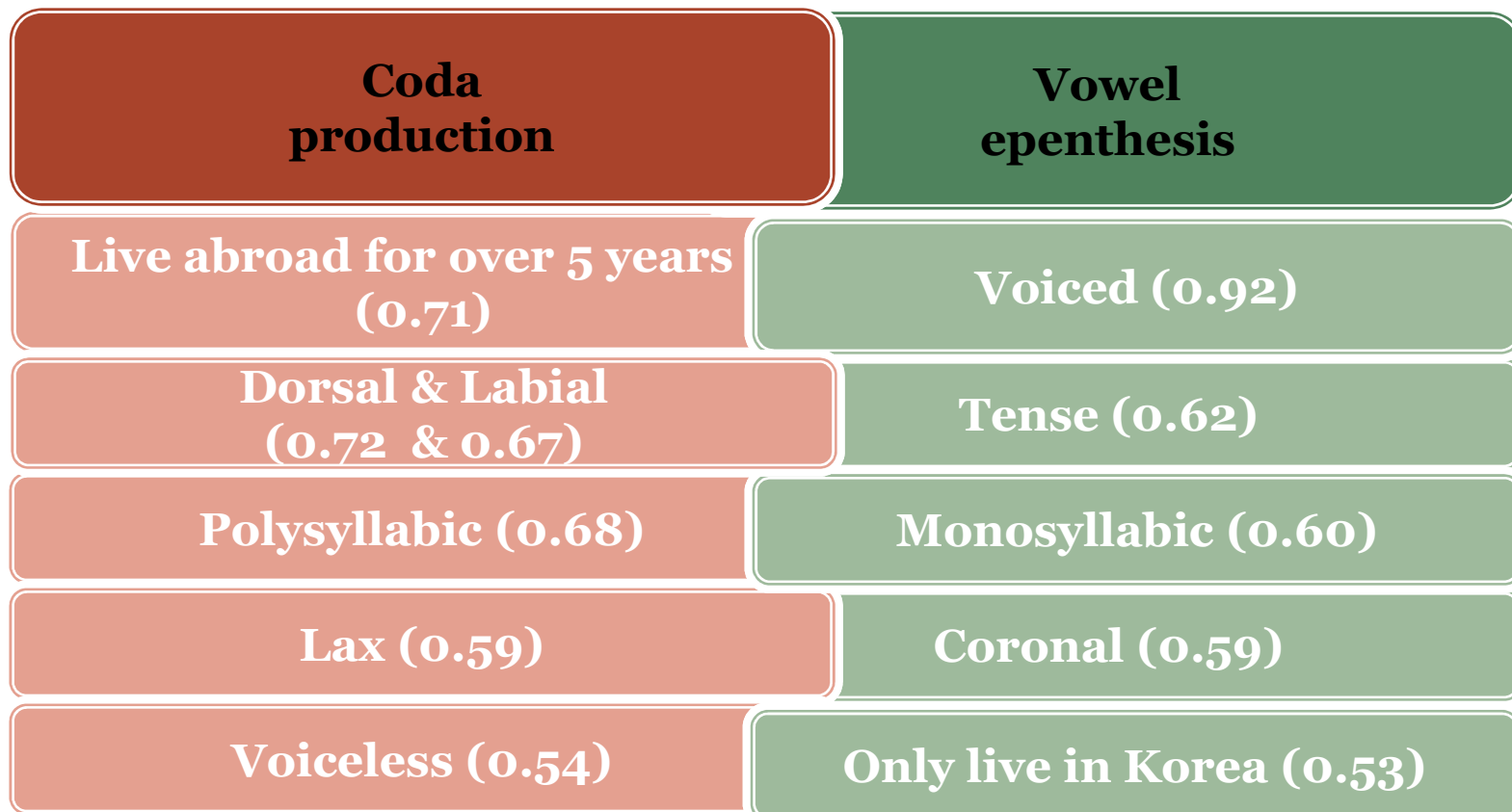
Interaction between length of residence & tenseness of the pre-final vowel
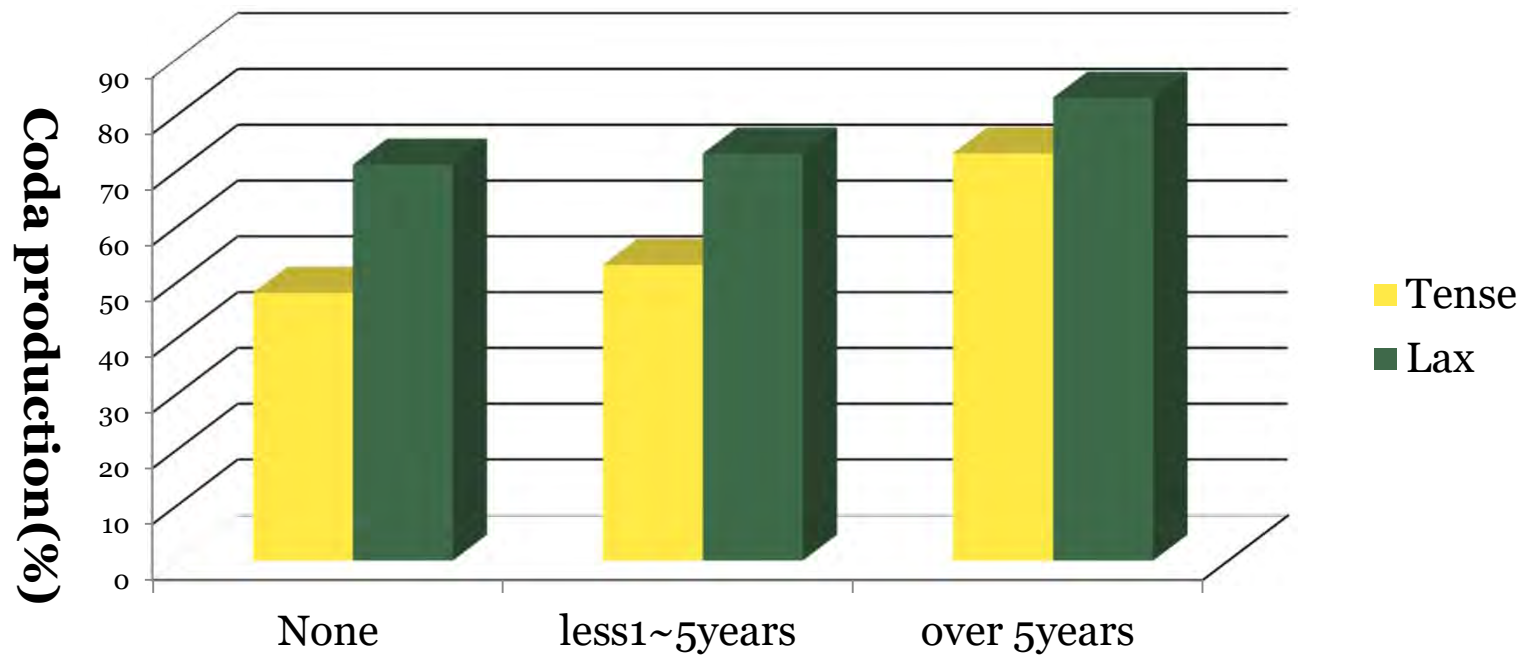

Figure 1 Coda production by length of residence & tenseness (%)

The occurrence of codas in English loanwords placed at the end in word-final stops increases in Koreans who have lived in English speaking countries for over five years.

Interaction between length of residence & place of articulation

Labial word-final consonants
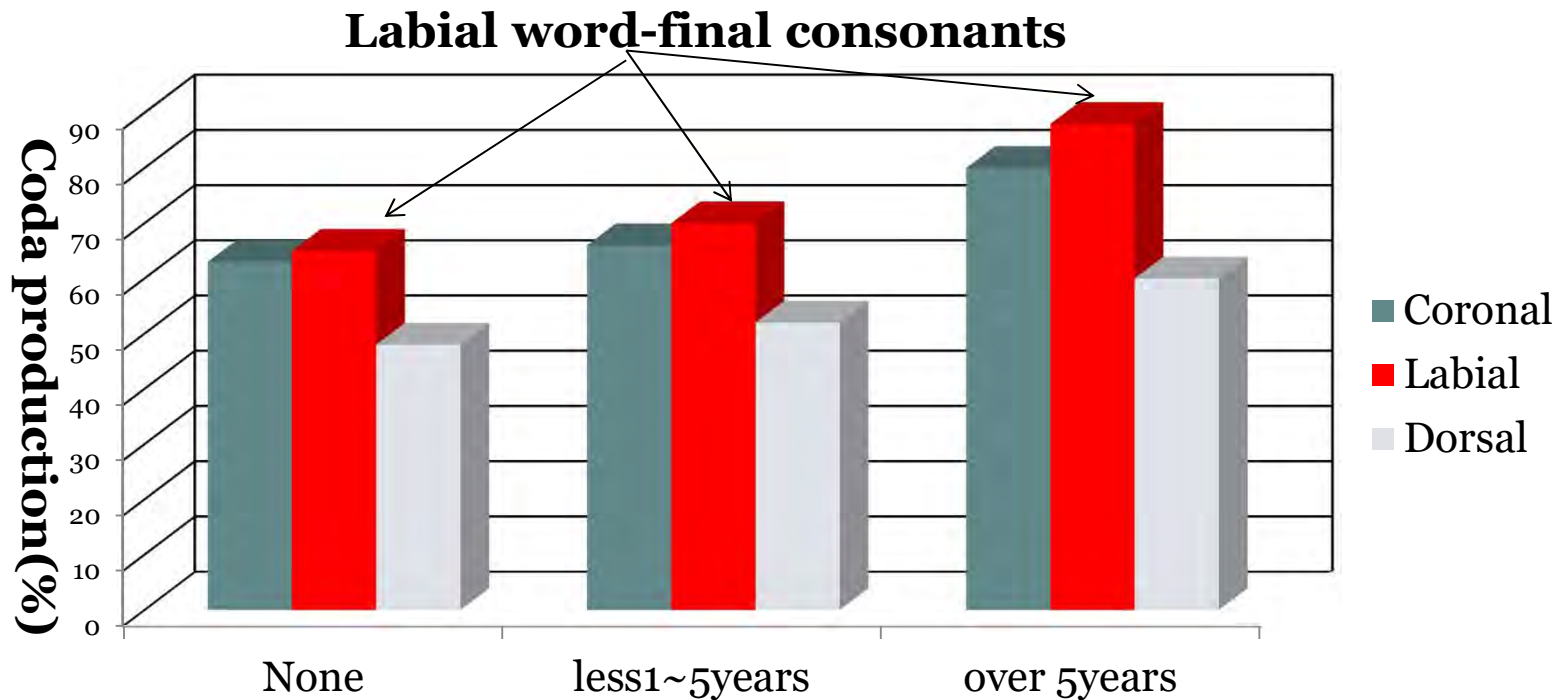
Figure 2 Coda production by length of residence & place of articulation (%)

Labial word-final consonants (88%) are more likely to surface as codas by Korean speakers who have stayed outside of Korea for over 5 years.

Interaction between length of residence & number of syllables
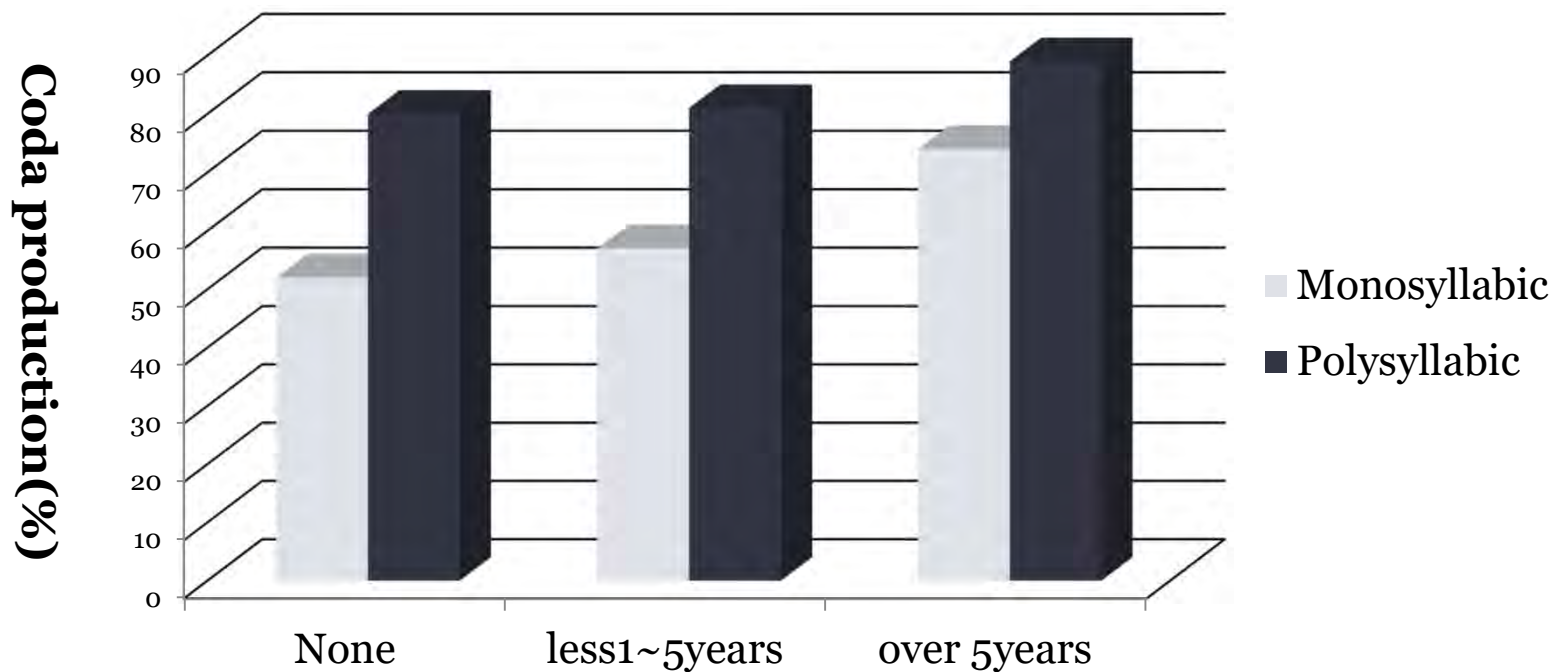
Figure 3 Coda production by length of residence & number of syllable (%)

When word size is more than two syllables, codas are more likely to appear in English word-final stops by Korean speakers.

Interaction between length of residence & Voicing of the final stop
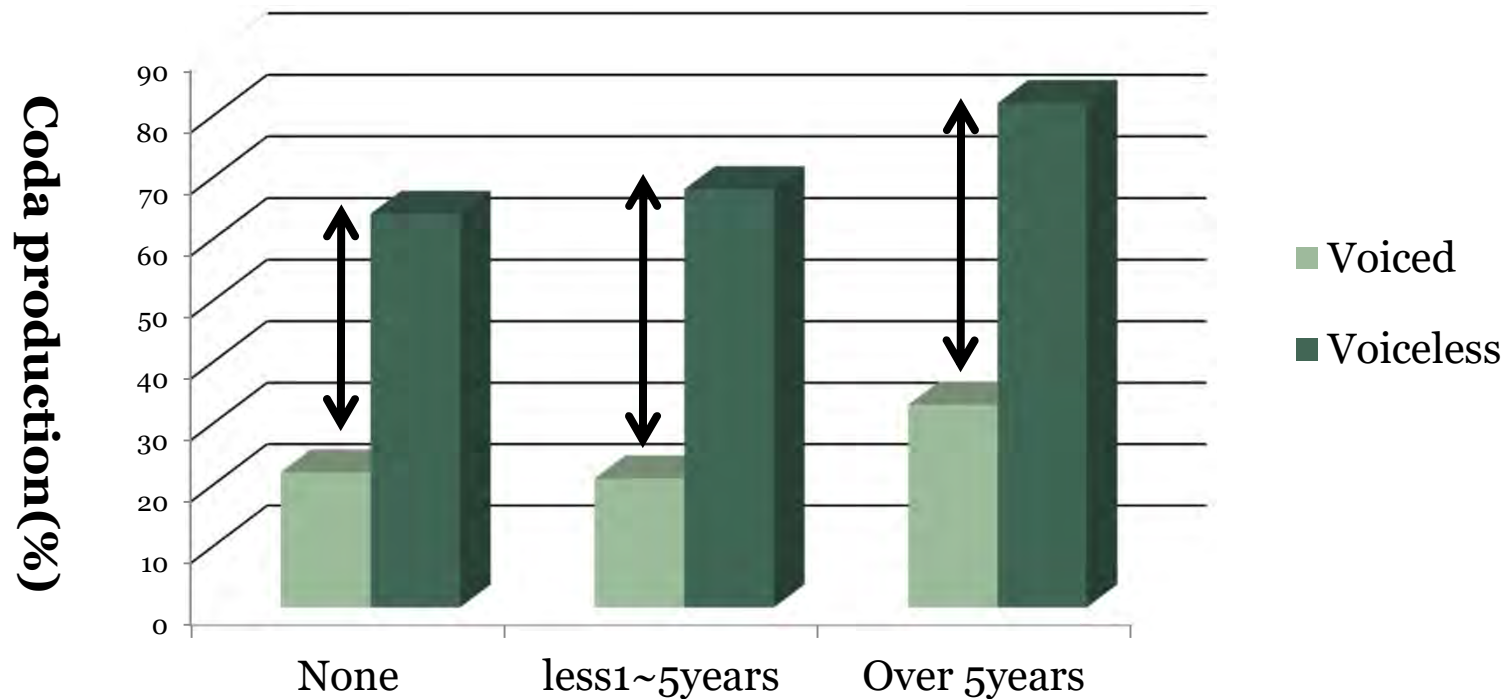
Figure 4 Coda production by length of residence & voicing of the final stop (%)

The possibility of coda production of English word-final stops by Koreans is higher when the stops are voiceless.

6. Stochastic OT analysis

- This paper adopts Boersma (1998) and Boersma & Hayes' (2001) methodology for investigating variability in the framework of Optimality Theory: **Stochastic OT(StOT).**

- Stochastic OT grammars in which every constraint has a ranking value along a continuous real-number scale .
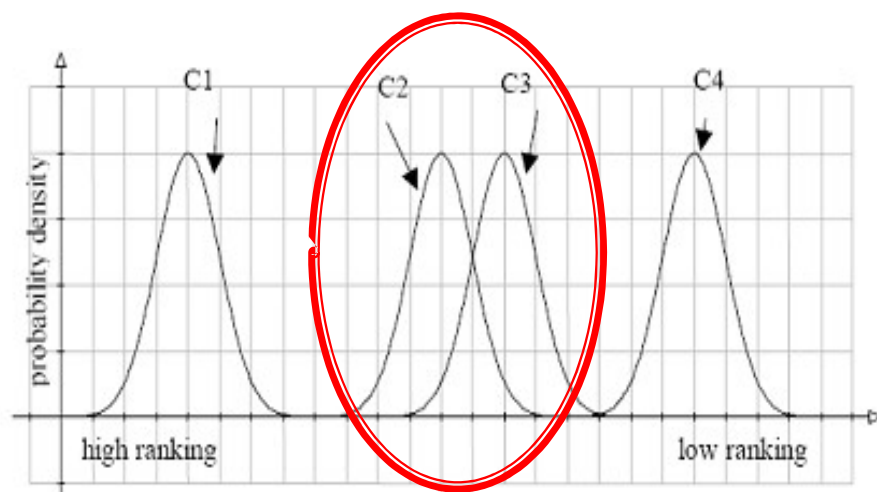
- We used Praat (Boersma & Weenink 2007) to determine the ranking values of the constraints.

Extra-linguistics/linguistic factors in StOT

One extra-linguistic factor

- **The length of residence**

Three different grammars

Level 1 / Level 2 / Level3

Proficiency levels correspond
to different grammars (Cordoso 2007)

Linguistic factors

- Tenseness of the pre-final vowel
- Voicing of the final stop
- Place of articulation
- The number of syllables

OT Constraints

- Ident(rel)
- NoCoda(vd) , NoCoda(vl)
- NoCoda(cor),NoCoda(dor),NoCoda(lab)
- WdMin

OT constraints

Constraints		
NoCoda(vd)	No voiced consonants in coda	(Boreslow et al 1998)
NoCoda(vl)	No voiceless consonants in coda	
NoCoda(cor)	No coronals in coda	
NoCoda(dor)	No dorsals in coda	
NoCoda(lab)	No labials in coda	
WdMin	Prosodic words are minimally disyllabic	(Bernhardt & Stemberger 1998, McCarthy & Prince 1995)
Dep	No epenthesis	
Ident(rel)	Obstruents in the output are identical in 'release' to obstruents in the input	Kang (2003)

Kang(2003) proposed that a stop tends to be released after a tense vowel in English, and this release is adapted with vowel epenthesis in Korean.

Constraints and ranking values (Level 1)

Level 1: ranking values

Constraint	Ranking value
Ident(rel)	101.971
Dep	101.549
NoCoda(cor)	101.124
NoCoda(vd)	100.944
WdMin	100.437
NoCoda(dor)	99.663
NoCoda(lab)	97.664
NoCoda(vl)	97.506

ex: 'rate'

	Coda occurrence	
	Observed(survey)	Produced(StOT)
Level 1	34.13	32.43

Tableau 1: Variable coda production in Level 1 (ex: 'rate')

/reit/	Ident (rel)	Dep	NoCoda (cor)	NoCoda (vd)	WdMin	NoCoda (dor)	NoCoda (lab)	NoCoda (vl)
[leit]	*!		*					*
→[leitɨ]		*						

Constraints and ranking values (Level 2)

Level 2: ranking values

Constraint	Ranking value
Dep	106.605
NoCoda(vd)	106.108
Ident(rel)	105.227
NoCoda(cor)	105.034
WdMin	104.656
NoCoda(dor)	99.605
NoCoda(lab)	88.756
NoCoda(vl)	87.136

ex: 'rate'

	Coda occurrence	
	Observed(survey)	Produced(StOT)
Level1	34.13	32.43
Level2	47.92	50.92

Tableau 2:Variable coda production in Level 2 (ex: 'rate')

/reit/	Dep	NoCoda (vd)	Ident (rel)	NoCoda (cor)	WdMin	NoCoda (dor)	NoCoda (lab)	NoCoda (vl)
→[leit]			*	*				*
[leitɨ]	*!							

Constraints and ranking values (Level 3)

Level 3: ranking values

Constraint	Ranking value
Dep	103.004
NoCoda(vd)	101.747
Ident(rel)	100.925
NoCoda(dor)	100.548
WdMin	99.560
NoCoda(cor)	99.673
NoCoda(lab)	96.775
NoCoda(vl)	95.249

ex: 'rate'

	Coda production	
	Observed(survey)	Produced(StOT)
Level1	34.13	32.43
Level2	47.92	50.92
Level3	66.67	67.80

Tableau 3: Variable coda production in Level 3(ex: 'rate')

/reit/	Dep	NoCoda (vd)	Ident (rel)	NoCoda (dor)	WdMin	NoCoda (cor)	NoCoda (lab)	NoCoda (vl)
→ [leit]			*			*		*
[leitɨ]	*!							

Grammars of three levels

Grammars by levels	Coda occurrence	
	Observed(all data)	Produced(all data)
Level 1	**64.48**	**63.72**
Ident(rel), Dep, NoCoda(cor) , NoCoda(vd) , WdMin, NoCoda(dor), NoCoda(lab), NoCoda(vl)		
Level 2	**68.25**	**63.42**
Dep, NoCoda(vd), Ident(rel), NoCoda(cor), WdMin, NoCoda(dor), NoCoda(lab), NoCoda(vl)		
Level 3	**80.74**	**81.63**
Dep, NoCoda(vd), Ident(rel) , NoCoda(dor), WdMin, NoCoda(cor), NoCoda(lab), NoCoda(vl)		

Evaluation noise (standard deviation): 2.0
Learning trials: 100,000
Initial state: 100
Plasticity: 0.1

7. Conclusion

- The GOLDVARB statistical results show that variation in the acquisition of English word-final codas by Korean speakers is triggered by linguistic and extra-linguistic factors.

- A Stochastic version of the framework of Optimality Theory accounts for the variable adaptations in English loanwords.

Thank you for your attention!

Any Questions?

	Level 1 (No experience)			
	Observed	produced(1st)	produced(2nd)	Prouduced(3rd)
Word	**Coda production (%)**	Coda production (%)	Coda production (%)	Coda production (%)
merit	**17.37**	74.73	83.84	70.24
cut	**58.38**	65.43	54.69	49.55
hip	**86.83**	81.03	60.48	62.32
boycott	**84.38**	74.44	83.69	70.03
jeep	**41.32**	35.29	34.10	34.70
target	**96.41**	74.74	83.83	70.09
weight	**25.15**	32.56	46.84	39.04
tag	**41.02**	63.55	44.59	42.16
dot	**52.99**	65.51	54.77	49.61
soup	**52.10**	35.36	34.04	34.79
carpet	**78.14**	74.53	83.73	70.16
format	**96.41**	74.46	83.80	69.95
internet	**97.60**	74.65	84.06	70.11
league	**3.89**	27.44	25.34	23.78
spot	**94.31**	74.68	84.01	70.30
bonnet	**73.05**	74.51	83.88	70.13
scout	**58.91**	32.39	46.99	39.25
cake	**49.70**	38.95	54.91	43.59
robot	**84.13**	74.33	83.74	70.18
rock	**97.60**	75.75	60.15	53.82
rate	**34.13**	32.43	47.01	39.26
site	**8.08**	32.67	47.14	39.41
net	**34.13**	65.23	54.65	49.77
goal_net	**62.57**	65.49	54.65	49.40
spirit	**77.25**	74.38	84.28	70.22
flute	**74.85**	32.46	46.47	39.21
pamphlet	**97.01**	74.56	83.84	70.42
tape	**46.11**	42.48	55.45	54.02
group	**98.20**	42.25	55.71	53.93
set	**23.35**	65.17	54.62	49.43

Evaluation noise:2.0, Plasticity 0.1 Trial per input 100000

Word	Level 2 (1~5 years)			
	Observed	Produced(1st)	Prouduced(2nd)	Prouduced(3rd)
	Coda production (%)	Coda production (%)	Coda production (%)	Coda production (%)
merit	23.61	78.98	70.53	83.54
cut	54.17	52.77	56.33	66.22
hip	86.11	60.23	71.85	75.11
boycott	81.56	79.30	70.56	83.46
jeep	46.15	40.47	51.58	32.24
target	96.53	79.08	70.73	83.50
weight	30.56	52.41	50.70	35.98
tag	39.58	45.65	46.25	61.02
dot	63.89	52.73	56.18	66.31
soup	60.42	40.81	51.84	32.28
carpet	75.69	79.09	70.79	83.53
format	97.22	78.80	70.69	83.48
internet	99.31	78.96	70.80	83.45
league	2.08	31.29	33.98	25.93
spot	95.14	79.03	70.69	83.53
bonnet	68.06	79.16	70.56	83.51
scout	62.94	52.45	50.83	36.01
cake	67.36	54.63	57.79	39.63
robot	86.81	79.07	70.51	83.52
rock	98.61	54.57	62.97	71.01
rate	47.92	52.42	50.92	36.03
site	9.03	52.50	50.90	36.04
net	37.50	52.64	56.34	66.23
goal_net	68.75	52.53	56.39	66.26
spirit	87.50	79.04	70.75	83.53
flute	79.17	52.56	50.68	36.05
pamphlet	97.22	78.97	70.29	83.52
tape	56.25	64.05	68.74	43.39
group	99.31	63.79	68.73	43.54
set	29.17	52.90	56.54	66.26
Evaluation noise:2.0, Plasticity 0.1 Trial per input 100000	Evaluation noise:2.0, Plasticity 0.1 Trial per input 100000			Plasticity 0.1 Trial 1000000

| Word | Observed | Produced(1st) | produced(2nd) | Prouduced(3rd) |
	Coda production (%)	Coda production (%)	Coda production (%)	Coda production (%)
			Level 3 (more than 5years)	
merit	**51.85**	92.03	82.57	83.22
cut	**55.56**	86.39	76.71	69.75
hip	**88.89**	92.31	88.27	78.25
boycott	**85.19**	91.85	82.81	83.40
jeep	**81.48**	71.60	72.31	58.67
target	**96.30**	91.89	82.62	83.25
weight	**66.67**	71.56	67.64	62.03
tag	**55.56**	63.78	63.46	54.93
dot	**81.48**	86.26	76.83	69.94
soup	**92.59**	71.65	72.11	58.51
carpet	**92.59**	91.95	82.36	83.41
format	**100.00**	91.88	82.56	83.36
internet	**100.00**	91.81	82.71	83.22
league	**11.11**	51.08	53.57	41.89
spot	**96.30**	91.97	82.49	83.40
bonnet	**85.19**	91.95	82.76	83.40
scout	**81.48**	71.48	67.77	61.79
cake	**74.07**	67.18	70.89	60.39
robot	**88.89**	91.94	82.47	83.33
rock	**100.00**	81.16	79.88	68.15
rate	**66.67**	71.51	67.80	61.56
site	**40.74**	71.40	67.83	61.67
net	**66.67**	86.35	76.83	69.87
goal_net	**81.48**	86.34	76.74	69.96
spirit	**92.59**	92.11	82.65	83.15
flute	**92.59**	71.18	67.64	61.77
pamphlet	**100.00**	91.94	82.42	83.40
tape	**77.78**	76.40	78.76	71.32
group	**100.00**	76.45	78.84	71.20
set	**66.67**	86.56	76.63	69.81
Evaluation noise:2.0, Plasticity 0.1 Trial per input 100000		Evaluation noise:2.0, Plasticity 0.1 Trial per input 100000		