

-

*

김미란** · 최재웅*** · 홍정하
(고려대학교)

Kim, Miran, Jae-Woong Choe and Jungha Hong. 2014. Distributional characteristics in Korean onset-nucleus sequences and hierarchical clustering of Korean vowels. *Studies in Phonetics, Phonology and Morphology* 20.1. 23-49. Native speakers' sensitivity to type and token frequency has been well recognized in many linguistic studies and is now known as frequency effects. In this study, we provide frequency information on the sequences of onset and nucleus by analyzing the complete lexical entries (510,032) listed in the Standard Korean Language Dictionary (2008). The frequency information has been extracted from a total of 1,761,214 syllables as they are represented in the dictionary. In addition to the frequency information, we provide a hierarchical cluster analysis of vowels referring to the frequency of individual onsets combined with individual vowels. Three main findings are reported in this paper. Firstly, diphthongs and monophthongs are clustered separately by hierarchical clustering with a characteristic that diphthongs are combined with a more restricted set of onsets when compared to monophthongs. Secondly, the two vowels 'ㅟ/ø/, ㅟ /y/, ' which are still being debated about regarding to whether they are monophthongs or diphthongs, are similar to a diphthong cluster with respect to the frequency information of combining onsets. Finally, the two vowels 'ㅚ/e/, ㅚ /æ/, ' whose surface forms are undergoing a merging process in modern Korean, show different sets of onsets that they combine with. The use of a large language resource allows us to explore empirically existing patterns, which may otherwise remain undiscovered or unpredicted by any theory that describes only the permissible combinations. **(Korea University)**

Keywords: hierarchical agglomerative cluster analysis, phoneme frequency, onset-nucleus sequence, frequency effects, monophthongs, diphthongs, Korean vowel system

1.

(frequency effect)

가
가

(Treiman et al. 2000, Albright 2002, Albright and

*

가

2013 4

** 1 , ***

Hayes 2002, Perruchet and Peereman 2004, Lee 2007a,b, Albright 2008, Jun 2010, Yun 2010, Lee 2011, Gries and Divjak 2012). Albright (2008) (morphological paradigm)
가

가
[]
(1989, Kang 2002).
[] +
.

(Jun 2010).

가
가 (word-likeness)
(-Lee 2007a,b, -Perruchet and
Peereman 2004, -Treiman et al. 2000).
'rup'-'ruk'('sif'-'hif')
가 'rup' 'ruk'
'sif'가 'hif'
(Treiman et al. 2000). /up/
/uk/ , /si/ /hi/

가

(/)
가 , 가
가
' + ' , 가
' + ' , 가
) 가 (+)
, 가 (+
/
,

가

,

.

/p, p^h, p^{*}/

(1) (2003: 327)⁴

1: / ()/
 (* , *)
 2: [+ , - l 가 /j/
 3: [- , + l 가 /w/
 (* , *)

가
 가
 가 (1)
 1 ~ 3
 가
 ,
 ,
 5
 (‘ㅇ’) , 21 가 , 30 가 (18 가
) . 11,340
 가 ,
 2,475 가 .⁶
 가 - -
 .

가 가

⁴
 가 가 가 . 가 4 ‘ㅡ’
⁵ 가 . (2003: 326) -
 가 / 가 가
 가
⁶ 11,172 (unicode) . 2,475
 . 10
 (: ‘ ’ -
 1,596
 , 10 2 544 , 1
 335 1,283 , 1,212
 (2011).
 - 가 가
 ,
 2,350 가 .

28 . .

가 ,

(1)

. ,

2.2

가

, (1) /세, H/ 가가
가, (2) /ㅌ,ㅍ/

가 7 10
가 7

1 .

1.

	8				
					10 : 1965, 1985, 1992, 1996, . 1997
	ㅣ	(ㅌ)	ㅡ	ㅌ	9 : 1997
	ㅅ	(ㅍ)	ㅍ	ㅍ	8 : 1996
	(H)		ㅌ		7 : 1988, 1993, 2000, 2003, 2003 .

10

.

10

(1 -).

가

, ‘H/세’가
10 가

⁷ , Kim (1968)
4 가

가

가

⁸ ‘ㅡ, ㅌ’
(2003)

(1) ‘ㅍ’ ‘ㅌ’
(1) , ‘ㅍ/ㅌ’ 8
(1)
(1)
가 7
‘ㄱ’ ‘ㅎ’ ,

9

가

‘ㅍ/ㅌ’가

‘ㅎ/ㄱ’

가

가

3.

가

10

3.1

:

(2008) .¹¹

510,032 ,

⁹ ‘ㅍ/ㅌ’ ‘ㄱ/ㅎ’ ‘ㄱ/ㅌ’가
, ‘ㅍ/ㅌ’
(1988, 2000, 2002, 2003,
2003, (2011).

¹⁰

(2008ㄱ,ㄴ, 2010)

¹¹

가

30

1,761,214¹² (2011)
 (47,401) 10 가
 1,540,376 , 1,761,214 , 2
 855,102
 , (2011)
 7
 2
 (2011: 209)

2.

(Type I)	(Type II)			
V	(V)	65,432: 3.72 % (3.2%)	102,529	5.82% (5.3%)
	(GV)	37,097: 2.11% (2.1%)		
CV	(CV)	725,055: 41.17% (42.8%)	803,583	45.63% (48.1%)
	(CGV)	78,528: 4.46% (5.3%)		
VC(C) ¹³	(VC)	67,689: 3.84% (2.9%)	118,309	6.72% (5.2%)
	(GVC)	50,620: 2.87% (2.3%)		
CVC(C)	(CVC)	647,954: 36.79% (36.1%)	736,793	41.83% (38.4%)
	(CGVC)	88,839: 5.04% (2.3%)		
			1,761,214	100%

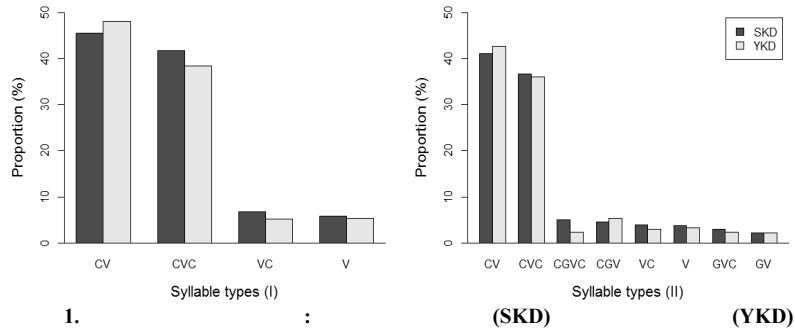
1 (2011)
 (8가),
 (4가) 1 2

가 ()

¹² 518,743 ‘ㄱ-ㄴ-ㄷ’
 (510,032). 4% 10%가
 (http://stdweb2.korean.go.kr/)

¹³ 2013 1)

‘CVC’
 (open syllable) (closed syllable)

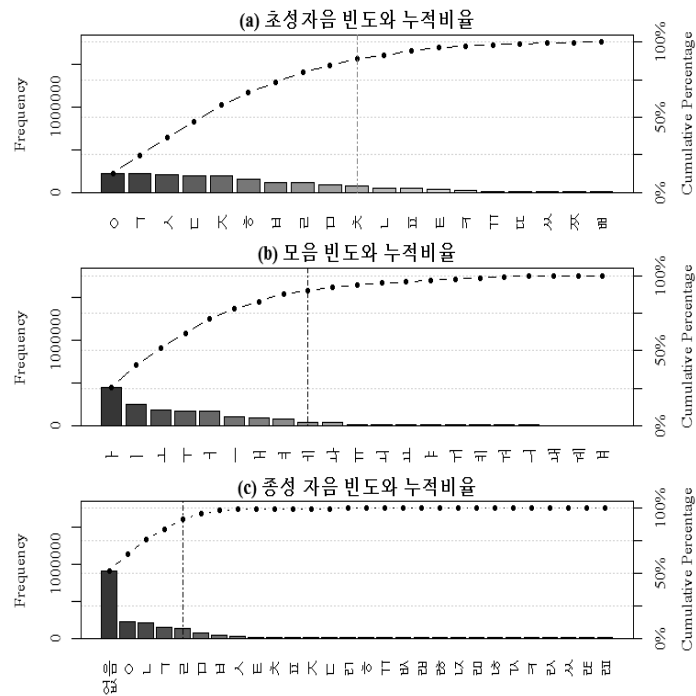


1. : (SKD) (YKD)
2 가 (Type I, II) 1

가
CGVC CGV
(2011) CGV(5.3%)가 CGVC(2.3%)
가 3% ,
CGV(4.46%)가 CGVC(5.04%)가 0.6%
가¹⁴ 가
CV(48.1%) , CVC (41.83%)
가
'(1)CV> (2) CVC > (3)
CGVC > (4) CGV > (5) VC > (6) GVC > (7) V > (8) GV'
(2011)
가 5%
¹⁵ 가
가 CGVC 2.7%
가¹⁶ , 1%

¹⁴ C(), V(), G()
¹⁵ (1) > (2) > (4) > (7) > (5) > (3) = (6) > (8) (2011)
(403,605)
'(1) CV(59.1%) > (2) CVC(21%) > (7) V(7.2%) > (8) GV(3.4%) > (4) CGV(3.1%)
> (5) VC(3.0%) > (3) CGVC(2.3%) > (6) GVC(0.9%)'가
(1) (2) 가
¹⁶ (2010, 2011)
가
(2010, 2011)
,

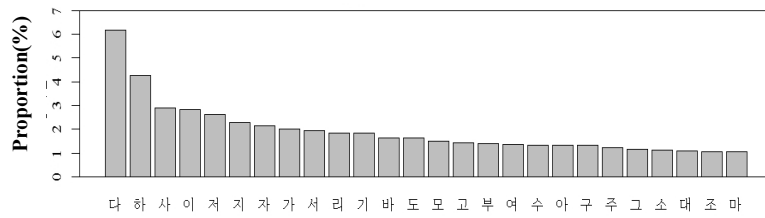
2 1,761,214 , ,
 (‘ㅇ’) (‘ㅇ’)가 ,
 (C)(G)V(C) ,
 12.5% , 51.4% .
 2b . (‘ㅏ, ㅣ, ㅑ, ㅓ, ㅕ...’ 가 가
 .)



2. (, ,)

2 가 90%
 (‘ㄱ, ㅅ, ㄷ, ㅈ’) 50% , ‘ㅏ, ㅣ,
 ㅑ’ 가
 50% ,
 ‘ㅇ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ’

90% , , 90%
가
가 26 (50.4%) 2 가



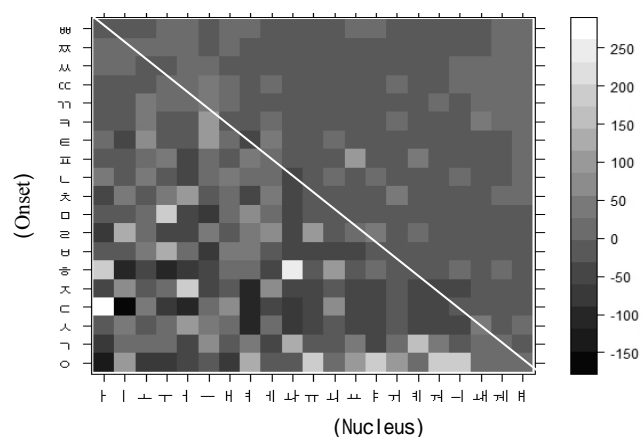
3. 50% (26 가)
3 가 ‘ (6.15%)’ ‘
(4.25%)’
> (2.29%) > (2.15%) > 가(2.01%) > (1.94%) > (1.85%)
(2.88%) > (2.83%) > (2.61%)
가

0.0025 (1/(19×21))
가 가 가
4 (Chi-Square) (residuals)
17 (2011: 204) > > > > >
(2011) , , , , , 7
1 , , (-)

(Heat map) .¹⁸

가
가
가
가⁴
가

19



가 가 ‘ㅏ’()
 , ‘ㅎ, ㄷ’
 , 가 ‘ㅎ’
 ‘ㅏ’ 가 가
 .
 가
 .
 .
 ,

18

19

holder)

‘O’

(place

가 (가) .

, .

3.2 (hierarchical agglomerative cluster analysis)

/

가

가 (Gries 2013). ,

가 . (bottom-up) (hierarchical agglomerative cluster analysis) . /

(dendrogram) .

, , (Dyen et al. 1992, Gries 2008, Hilpert and Gries 2009, 2010, Gries and Otani 2010).

/ 'Canberra' , (linkage) 'Ward' ²⁰ R(R Core Team 2014: 3.0.3) ²¹ .

가

가

가 , 가 , 가 3a .

²⁰ Gries (2008) Gries and Otani (2010) . R

²¹ <http://cran.r-project.org/>.

Gries (2013) . R

가 /

Canberra , Ward

(Gries 2008). Canberra $\sum(|x_i - y_i| / |x_i + y_i|)$

3c가 가 .

3.

	(a)	(b)				(V-prop)						(c)	(C-prop)			
	중성(모음)															
		ㅏ	ㅓ	...		ㅏ	ㅓ	ㅗ	...			ㅏ	ㅓ	...		
초	ㄱ	35514	16641	...	ㄱ	0.165	0.077	0.118	...	100%		ㄱ	0.011	0.552	...	
성	ㄴ	15694	2245	...	ㄴ	0.322	0.046	0.169	...	100%		ㄴ	0.026	0.003	...	
자	ㄷ	108581	6272	...	ㄷ	0.561	0.032	0.147	...	100%		ㄷ	0.002	0.002	...	
음	ㄹ	14974	6537	...	ㄹ	0.131	0.057	0.011	...	100%		ㄹ	0.172	0.091	...	
	ㅁ	18643	4004	...	ㅁ	0.205	0.044	0.131	...	100%		ㅁ	0.001	0.001	...	
	ㅂ	28530	13318	...	ㅂ	0.242	0.113	0.136	...	100%		ㅂ	0.004	0.000	...	
	
												...	100%	100%	100%	

		3b		
A	B		가	가
.			가	
.	3a-c가			
.	3a	가		
	‘ㄷ+ㅏ (108,581) > ㄱ+ㅏ (35,514) > ㅁ+ㅏ (28,530)...’	(V-prop)	.	3b
			‘ㅏ (16.56%) > ㅑ	가
	‘ㄱ,’		가	,
(11.84%) > ㅓ (7.76%) > ...’				가
‘ㄱ, ㄹ, ㅁ, ㅂ,’				
	‘ㅏ,’ ‘ㅑ,’			
	‘ㅓ,’ ‘ㅕ,’		가	
.				
	가		3c	
	(C-prop)		,	
‘ㅏ,’	‘ㄹ > ㄴ > ㄱ...’		가	
‘ㅓ,’	‘ㄱ > ㄹ > ㄴ >...’	,	‘ㅑ,’	‘ㄱ
> ㄹ > ㅁ > ...’	가	.		
가	가		가	

22
() ()

가

가

, 가 가

, 3c /

4.

가

가

가

22

23

가

(phasing) , (in-phased) 가
(anti-phased)
(Browman and Goldstein 1989, 2000).
() 가

(assimilation)

, (' > ')
, ' 가
(: ' , , ' > ' , , ,
, > , 가 > 가 , , ,).

가

38

4.1

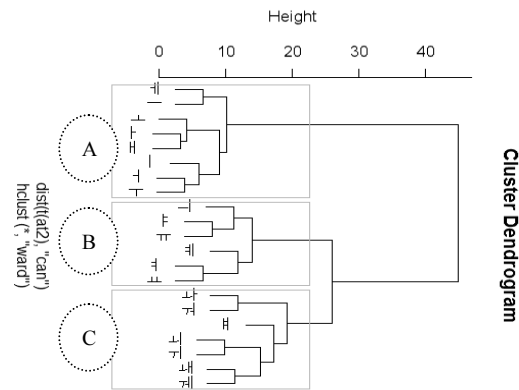
3.1

(5)

19

21

(C-prop)



5

‘Canberra, Ward’

가

5

가

가 가
(height=0)

가 , ‘ㄱ/ㄷ’

가 ()

가 가
가 4

가 2
가 2

가

가

5

(A, B, C)

가
‘ㄱ/ㄷ’

8

10

‘ㄱ, ㅌ, ㅍ, ㅊ, ㅋ, ㆁ’ 6

j-

²⁴ ‘ㅈ’
 ‘ㅉ, ㅊ, ㅌ, ㅍ, ㅑ, ㅓ’가

²⁵

가
 (B, C)

²⁶

(1)

(8)

‘ㅌ/ㅍ’ j- , w-
 가 ‘ㅌ/ㅍ’가

²⁴ ‘ㄹ’ (off-glide) (on-glide) ‘ㄴ’
 가 , 가 (2011) ‘ㄹ’
 [w i] ‘j-’ ‘w-’

²⁵ ‘ㅈ’ 1,761,214 ‘ㄱ(5), ㄴ(13),
 ㄷ(23), ㄹ(1), ㅈ(2)’
 가 (: ‘ ()’, ‘- (‘- , ,)’, ‘ (‘- ,)’.

²⁶ (linearity)
 가 1) (: ‘ = /nj+a/’) ,
 2) (: ‘ = /n+ja/’)
 가 ,
 가 () 가 ,
 (가) 가

4.2

8, (2) ‘ㅍ/ㅑ’, (3) j-, (4) w- (1)

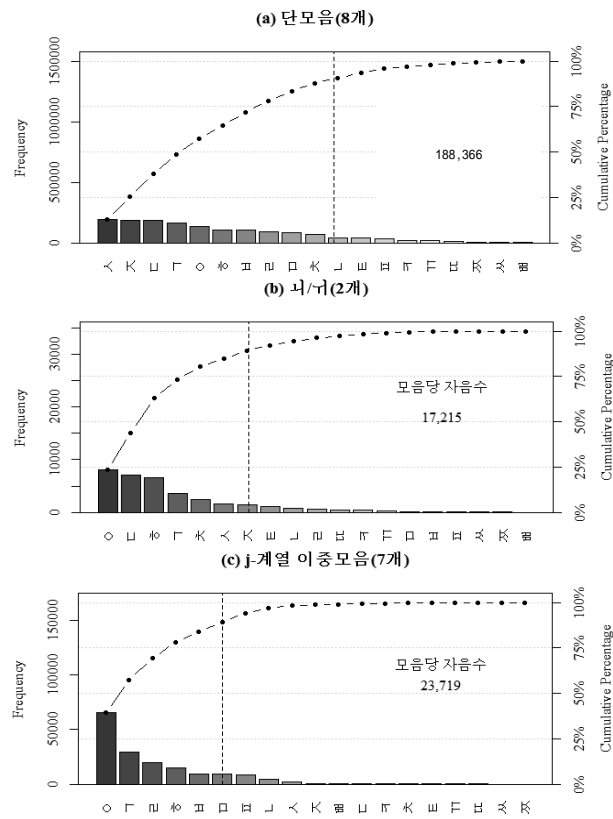
, 17,215 (‘ㅍ/ㅑ’), 23,719 (j-), 13,654 (w-)
가

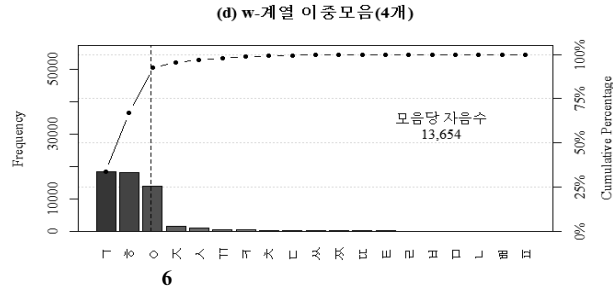
6

0~100%

가

90%





6

90%

10

(c-d) 2~5

90%

(‘ㅓ’) 가

‘ㅓ/ㅓ’

가 가

27

‘ㅓ/ㅓ’

가 188,366

10 가

‘ㅓ/ㅓ’ (17,215)

(23,719; 13,654)

‘ㅓ/ㅓ’

‘ㅓ/ㅓ’

‘ㅓ/ㅓ’

7

‘ㅓ/ㅓ’

27

‘ㅓ(20,640)’, ‘ㅓ(13,827)’

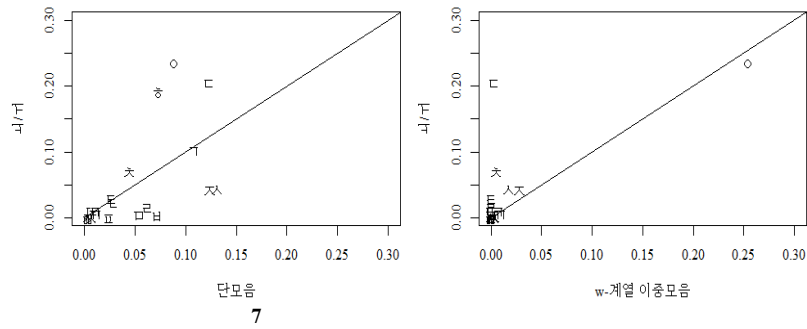
1.95%

가

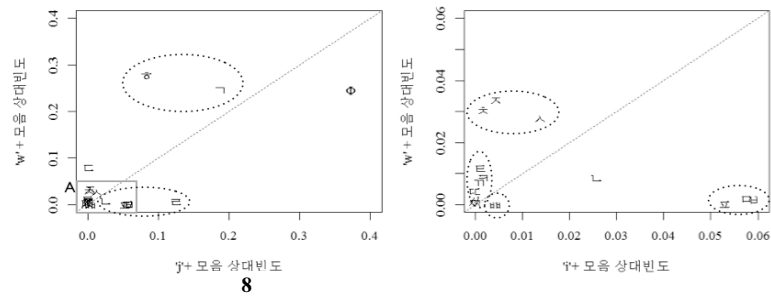
‘ㅓ/ㅓ’가

‘ㅓ/ㅓ’가

(), ‘ㅍ/ㅑ’ w-
 ()
 가 가
 가 가
 가



7
 ‘ㅏ’ ‘ㅑ/ㅑ’ 가 , ‘ㅓ’, ‘ㅑ’,
 ㅓ, ㅑ, ㅑ, ㅑ ‘ㅑ/ㅑ’ 가
 가 ‘ㅑ/ㅑ’
 ‘ㅑ/ㅑ’ w-
 (, 가 ‘ㅑ’) ‘ㅑ/
 ㅑ’ ‘ㅑ/ㅑ’ w-
 (5)
 가
 ()
 가 j- w-



8 w- ‘흥, ㄱ, (ㄷ)’
, (‘Φ’^{j-})^{j-}
가 가
.
‘ㅍ, ㅁ, ㅂ’ w-
(: ‘人, ㅈ, ㅊ’ ‘ㅌ, ㅋ, ㆁ, ㄷ’) j-
.

4.3 / 𐌺 / / 𐌿 /

가

28.

,

가 가

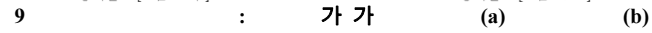
5

‘—’, ‘T’

‘—/T’

(

28 ‘ \mathbb{H} ’, ‘ \mathbb{H} ’, 19
(Lee and Ramsey 2011).



) 12

‘T’
‘T’

9b

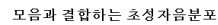
‘—/|’

‘ H / -H ’

(5)

 $\frac{1}{2} \text{H} / \text{H}$

10



10 ‘H / H’

, 가

‘ H ’

‘ \neg , \exists , $\bar{\sigma}$,

□'

가

(○), 天, 己, 天,

가

가

46,837 ‘ㅁ’가 ‘ㅂ’
‘ㅁ’
(‘ㅈ, ㄹ, ㅊ’) 4, 7, 9
가
‘ㅂ’
가
30
‘ㅁ’가 90,302 ‘ㅂ’가
‘ㄱ, ㄴ, ㅎ, ㅁ’
(ㅇ)’
‘ㅂ’
1, 3, 5, 6
가
29
‘ㅁ’가
,
가

6.

가
.
,
가
가
8
90% 10~11
가
5~6
(6).
,
.
‘ㅂ/ㅈ’가
가
.
‘ㅂ/
ㅈ’
.
‘ㅂ/ㅈ’
29
가
‘ㅂ/ㅈ’
,
‘ㅁ/ㅂ’
가
가
.
가
30
가 ()
‘가 가’
/

- . 1992.
- . 1996.
- . 2003.
- . 2000.
- _____. 2008¹. 13.2, 193-215.
- _____. 2008². 13.3, 477-502.
- _____. 2010. 15.1, 94-106.
- _____. 2011.
- . 2003.
- . 1997.
- . 1988. 12, 61-72.
- . 1997.
- . 1993. , 23-38. 3.1,
- . 1996.
- . 2003.
- . 24.
- . 2010.
- . 20, 127-160.
- . 2002.
- . 1965.
- _____. 1985.
- ALBRIGHT, ADAM. 2002. *The Identification of Bases in Morphological Paradigms*. PhD Dissertation. UCLA.
- _____. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.). *Language Universals and Language Change*, 144-181. Oxford University Press.
- ALBRIGHT, ADAM and BRUCE HAYES. 2002. Modeling English past tense intuitions with minimal generalization. *SIGPHON 6: Proceedings of the 6th Meeting of the ACL Special Interest Group in Computational Phonology*, 58-69. Philadelphia.
- BROWMAN, CATHERINE and LOUIS GOLDSTEIN. 1989. Articulatory gestures as phonological units. *Phonology* 6, 201-251.
- _____. 2000. Competing constraints on intergestural coordination and self-organization of phonological structure. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée* 5, 25-34.
- DYEN, ISIDORE, JOSH PH. B. KRUSKAL and PAUL BLACK. 1992. An

- Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82.5, 1-132.
- GRIES, STEFAN TH. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13.4, 403-437.
- _____. (2013) *Statistics for Linguistics with R* (2nd revised edition). De Gruyter Mouton. [] (). 2013.
- _____. :R .]
- GRIES, STEFAN TH. and DAGMAR DIVJAK. 2012. *Frequency Effects in Language Learning and Processing*. Trends in Linguistics Volume 1. De Gruyter Mouton.
- GRIES, STEFAN TH. and NAOKI OTANI. 2010. Behavioral profiles: a corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34, 121-150.
- HILPERT, MARTIN and STEFAN TH. GRIES. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24.4, 385-401.
- JUN, JONGHO. 2010. Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19.2, 137-197.
- KANG, YUNJUNG. 2002. Sound changes affecting noun-final coronal obstruents in Korean. Paper presented at *the 12th Japanese/Korean Linguistics Conference*. CUNY.
- KIM, CHIN-WU. 1968. The vowel system of Korean. *Language* 44.3, 516-527.
- LEE, KI-MOON and ROBERT S. RAMSEY. 2011. *A History of the Korean Language*. Cambridge University Press.
- LEE, YONGEUN. 2007a. Effects of Inter-phoneme probabilities on the acceptability judgment of Korean CVC nonwords. *Speech Sciences* 14.4, 41-52.
- _____. 2007b. English speakers' sensitivity to the phonotactic probabilities of 'CV-' and '-VC' components of CVC syllables in English. *Studies in Phonetics, Phonology and Morphology* 13, 105-121.
- _____. 2011. The gradient nature of syllable affiliations of Korean glides. *Korean Journal of Linguistics* 36.3, 735-764.
- PERRUCHET, PIERRE and RONALD PEEREMAN. 2004. The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics* 17, 97-119.
- R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org> (R version 3.0.3).
- SOHN, HO-MIN. 1999. *The Korean Language*. Cambridge University Press.
- TREIMAN, REBECCA, BRETT KESSLER, STEPHANIE KNEWASSER, RUTH

TINCOFF and MARGO BOWMAN. 2000. English Speakers' Sensitivity to Phonotactic Patterns, In Janet Pierrehumbert and Michael Broe (eds.). *Laboratory Phonology V: Acquisition and Lexicon*, 269-282. Cambridge University Press.

YUN, GWANHI. 2010. The effects of neighborhood density and word frequency on the production of English-learning Korean speakers. *Studies in Phonetics, Phonology and Morphology* 16.2, 217-243.

136-701 145

e-mail: mirany@korea.ac.kr

136-701 145

e-mail: jchoe@korea.ac.kr

136-701 145

e-mail: kleist@korea.ac.kr

received: February 26, 2014

revised: April 8, 2014

accepted: April 10, 2014