

Neutralization and anti-homophony in Korean

Daniel Silverman



Abstract. Neutralization in Korean involves a large number of oppositions, and affects a significant portion of the lexicon. Nonetheless, it induces remarkably little homophony. I propose herein that these highly divergent facts may be related: the neutralization of so many values in Korean is tolerated exactly because it has a negligible effect on the amount of derived homophony. Understanding how this came to pass requires an investigation of the history of Korean, especially in the context of its extensive contact with Middle Chinese. The present findings suggest a reconsideration of the role that ‘functional load’ (Martinet 1952, Hockett 1967) plays in patterns of sound change.

...There is no doubt that in some way or other, linguistic systems respond to change in ways that maintain meaning—more or less -William Labov (1994:569)

- Silverman (2006): a primary origin of allophonic (i.e. non-neutralizing) alternation is selectional pressures acting on sound systems: perceptually unambiguous speech tokens are more likely perceived with the meaning intended by speakers, and so are more likely to be reproduced as these listeners become speakers (see also Wedel 2006).
- The obvious question: whence neutralization and merger?
- According to Martinet (1952), the tendency toward merger of an opposition is favored to the extent that
 - 1) The values in opposition are phonetically similar
 - 2) The number of minimal morpheme pairs that the opposition is responsible for is low
 - 3) The number of minimal pairs within a correlated opposition is low (or the opposition is uncorrelated; “correlated”: a series that is opposed to another by one feature (Trubetskoy 1939))
 - 4) The minimal pairs belong to different syntactic categories
 - 5) The token frequency of one or both members of the minimal pairs is low
 - 6) The presence of additional morphological markers serves a disambiguating function
- If sounds are to merge, the result—*across the lexicon in use*—should not yield a significant increase in the amount of communicative confusion.
- The proposal: neutralizing alternations are more likely arise in a language if they do not significantly increase the level of homophony: ***derived homophony is not excessive.***
- The wording of this hypothesis is intentionally vague, because I do not pretend to be able to zero in on a numerical upper limit of derived homophony

Neutralization in Korean

- In Korean, many patterns of neutralization are tolerated and continue to expand because derived homophony is remarkably meager.
- Source for this study: The Sejong Project, a database of 1.5 million words gathered from written sources (<http://sejong.or.kr/>).
- Seven routes to C+C / C# neutralization (Martin 1992):

(A) Aplosivization

(B) Nasal lateralization

(C) Liquid nasalization

(D) Nasal assimilation

(E) Coronal assibilation

(F) Cluster reinforcement

(G) Coronal assimilation (variable)

- (1) Korean intervocalic values (shaded cells are variably implemented; adapted from Martin 1992)

$\begin{smallmatrix} +C \rightarrow \\ C \rightarrow \end{smallmatrix}$	p	p ^h	p'	t	t ^h	t'	s	s'	tʃ	tʃ ^h	tʃ'	k	k ^h	k'	m	n	ɾ	h	Ø
b	p'	p ^h	p'	pt'	pt ^h	pt'	ps'	ps'	ptʃ'	ptʃ ^h	ptʃ'	k'	k ^h	k'	m:	mn	mn	p ^h	b
p ^h	p'	p ^h	p'	pt'	pt ^h	pt'	ps'	ps'	ptʃ'	ptʃ ^h	ptʃ'	k'	k ^h	k'	m:	mn	mn	p ^h	p ^h
d	p'	p ^h	p'	t'	t ^h	t'	s'	s'	tʃ'	tʃ ^h	tʃ'	k'	k ^h	k'	m:	ɾ:	ɾ:	t ^h	d
t ^h	p'	p ^h	p'	t'	t ^h	t'	s'	s'	tʃ'	tʃ ^h	tʃ'	k'	k ^h	k'	m:	ɾ:	ɾ:	t ^h	t ^h
s	p'	p ^h	p'	t'	t ^h	t'	s'	s'	tʃ'	tʃ ^h	tʃ'	k'	k ^h	k'	m:	ɾ:	ɾ:	s'	s
s'	p'	p ^h	p'	t'	t ^h	t'	s'	s'	tʃ'	tʃ ^h	tʃ'	k'	k ^h	k'	m:	ɾ:	ɾ:	s'	ʃ'
dʒ	p'	p ^h	p'	t'	t ^h	t'	s'	s'	tʃ'	tʃ ^h	tʃ'	k'	k ^h	k'	m:	ɾ:	ɾ:	tʃ ^h	dʒ
tʃ ^h	p'	p ^h	p'	t'	t ^h	t'	s'	s'	tʃ'	tʃ ^h	tʃ'	k'	k ^h	k'	m:	ɾ:	ɾ:	t ^h	tʃ ^h
g	kp'	kp ^h	kp'	kt'	kt ^h	kt'	ks'	ks'	ktʃ'	ktʃ ^h	ktʃ'	k'	k ^h	k'	ɳm	ɳn	ɳn	k ^h	g
k ^h	kp ^h	kp ^h	kp ^h	kt'	kt ^h	kt'	ks'	ks'	ktʃ'	ktʃ ^h	ktʃ'	k'	k ^h	k'	ɳm	ɳn	ɳn	k ^h	k ^h
k'	kp'	kp ^h	kp'	kt'	kt ^h	kt'	ks'	ks'	ktʃ'	ktʃ ^h	ktʃ'	k'	k ^h	k'	ɳm	ɳn	ɳn	k ^h	k'
m	mb	mp ^h	mb	md	mt ^h	mt'	ms'	ms'	mtʃ'	mtʃ ^h	mtʃ'	ɳk'	ɳk ^h	ɳk'	m:	mn	mn	mh	m
n	mb	mp ^h	mp'	nd	nt ^h	nt'	ns'	ns'	ntʃ'	ntʃ ^h	ntʃ'	ɳk'	ɳk ^h	ɳk'	m:	ɾ:	ɾ:	nh	n
ŋ	ɳb	ɳp ^h	ɳp'	ɳg	ɳt ^h	ɳt'	ɳs'	ɳs'	ɳtʃ'	ɳtʃ ^h	ɳtʃ'	ɳk'	ɳk ^h	ɳk'	ɳm	ɳn	ɳl	ɳh	ŋ
l	lb	lp ^h	lp'	ld	lt ^h	lt'	ls'	ls'	ltʃ'	ltʃ ^h	ltʃ'	lg	lk ^h	lk'	lm	l:	l:	lh	l
Ø	b	p ^h	p'	d	t ^h	t'	s'	s'	dʒ	tʃ ^h	tʃ'	g	k ^h	k'	m	n	ɾ	Ø	Ø

- 304 possible combinations reduce to 75 values

(A) Aplosivization induces very little homophony

- Laryngeal neutralization is quite prevalent among non-prevocalic stops, and virtually unattested among prevocalic ones (Lombardi 1991, Steriade 1995, 1997, 2000). This position of neutralization typically involves *the loss of stop release*, or *aplosivization*.
- For aerodynamic and auditory reasons, stop releases are the optimal location for laryngeally-based cues (Kingston 1985, 1990, Bladon 1986, Silverman 1995, 1996, Wright 2004).

- Non-prevocally, the perceptual distinction among contrastive laryngeal states may be extinguished completely. This is laryngeal neutralization due to aplosivization.

(2) Alternation with aplosives (shaded values don't alternate; they never appear morpheme-finally.)

	Lexically prevocalic				Alternate with	Lexically non-prevocalic
	<u>Plain</u>	<u>Voiced</u>	<u>Aspirated</u>	<u>Tensed</u>		<u>Aplosive</u>
Labial:	p	b	p^h	p'		p'
Coronal:	t	d	t^h	t'		t'
	tʃ	dʒ	tʃ^h	tʃ'		
	s		(h)	s'		
Dorsal:	k	g	k^h	k'		k'

(3) Examples of neutralization due to aplosivization (from Jun 2007)

	Plosive:		Aplosive:	Gloss:
	(a) Locative (-e)	(b) Nominative (-i)		
Labial:	pab-e	pab-i	pap'	rice
	ip^h-e	ip^h-i	ip'	leaf
Coronal:	os-e	oʃ-i	ot'	clothes
	pat^h-e	patʃ^h-i	pat'	field
	nadʒ-e	nadʒ-i	nat'	day
	pitʃ^h-e	pitʃ^h-i	pit'	light
Dorsal:	kug-e	kug-i	kuk'	soup
	puək^h-e	puək^h-i	puək'	kitchen
	pak'-e	pak'-i	pak'	outside

- The Sejong Project corpus lists 34,803 different nouns.
- 10,138 nouns possess aplosive alternants
- 5,299 are word-final
- 4,839 are word-internal

(4) Distribution of word-final (#) and morpheme-final (+) obstruents/h for 10,138 nouns.

<u>Labials</u>		<u>Coronals</u>		<u>Dorsals</u>	
Non-neutralized value	Number of words	Non-neutralized value	Number of words	Non-neutralized value	Number of words
b #	1,154	d #	3	g #	3,522
b +	762	d +	0	g +	3,272
p^h #	189	t^h #	63	k^h #	12
p^h +	33	t^h +	60	k^h +	1
p' #	0	t' #	0	k' #	15
p' +	0	t' +	0	k' +	8
		d₃ #	10		
		d₃ +	41		
		tʃ^h #	74		
		tʃ^h +	43		
		tʃ' #	0		
		tʃ' +	0		
		s #	257		
		s +	612		
		(h) #	0		
		(h) +	5		
		s' #	0		
		s' +	2		
p^ʔ #	1,343	t^ʔ #	407	k^ʔ #	3,549
p^ʔ +	795	t^ʔ +	763	k^ʔ +	3,281
Total:	2,138	Total:	1,170	Total:	6,830
5299 nouns possess word-final neutralized aplosive alternants; 4839 nouns possess neutralized word-internal aplosive alternants; 10,138 out of 34,803 nouns; 29% of all nouns					

- 19.2% consists of **g**-final elements, and there are only 35 words with which these words might alternate.
- 6% are **b**-final, which potentially neutralize with only 72 other forms.

(5) Examples of homophony due to aplosivization

na d₃ -i	day (subject)	na t^ʔ -k'-wa	day and
na tʃ^h -i	face (subject)	na t^ʔ -k'-wa	face and
na f -i	sickle (subject)	na t^ʔ -k'-wa	sickle and

- The Sejong Project corpus tallies *inflected words across the lexicon in use*, and not simply bare roots.
- It is not the case that *all* potential derived homophonic *roots* are documented here.

(6) Exhaustive list of homophones due to aplosivization (throughout, shaded sets are false positives due to spelling errors)

Set number	Homophonous words	Non-neutralized allomorphs	Word number	Hangul	Token count	Gloss
1	tʃʌt̚	tʃʌdʒ	1	젓	44	breast/milk
		tʃʌs	2	젓	5	salted fish
2	tʃʌtkarak̚	tʃʌs-ka-rag	3	젓가락	27	chopsticks
		tʃʌdʒ-ka-rag	4	젓가락	3	unknown
3	tʃip̚	tʃib	5	집	9	house
		tʃip̚h	6	짚	19	straw
4	tʃot̚	tʃodʒ	7	좇	21	penis
		tʃot̚f̚h	8	좇	2	unknown
5	ip̚	ib	9	입	1,139	mouth
		ip̚h	10	잎	158	leaf
6	kʌt̚	kʌs	11	것	40,544	one
		kʌt̚h	12	겉	120	surface
7	mat̚	mas	13	맛	392	taste
		mat̚h	14	말	3	nearby place
8	miltʃip̚	mil-tʃip̚	15	밀집	21	crowd
		mil-tʃip̚h	16	밀짚	2	straw
9	nat̚	nadʒ	17	낮	294	day
		nat̚f̚h	18	낫	91	face
		nas	19	낫	25	sickle
10	odʒirap̚	o-tʃi-rab	20	오지랍	4	unknown
		o-tʃi-rap̚h	21	오지랴	3	front of outer garment
11	pak̚	pag	22	박	581	gourd
		pak̚	23	밖	1,568	exterior
12	pat̚	pat̚h	24	밭	380	heritage
		pad	25	밭	2	field
13	pit̚	pit̚f̚h	26	빛	726	light
		pidʒ	27	빚	128	debt
		pis	28	빗	6	comb
14	pʲʌt̚	pʲʌt̚h	29	별	32	sunshine
		pʲʌs	30	벧	6	crest (of fowl)
15	sɛudʒʌt̚	sɛ-u-tʃʌs	31	새우젓	8	unknown
		sɛ-u-tʃʌdʒ	32	새우젓	2	salted shrimp
16	sut̚	sut̚f̚h	33	숯	14	charcoal
		sut̚h	34	술	7	hair density

17	tʌt̚ʼ	tʌt̚ʰ	35	덜	31	trap
		tʌs	36	덜	4	a short time
18	tʰe-ip̚ʼ	tʰe-ip̚ʰ	37	테잎	2	tape
		tʰe-ib̚	38	테입	9	<i>unknown</i>
19	tʃipt̚ʼan	tʃib̚-tan	39	집단	400	group
		tʃip̚ʰ-tan	40	짚단	12	sheaf of straw
20	nat̚ʃʼak̚ʼ	nat̚ʃʰ-tʃʼag	41	낯짝	5	face
		nat̚ʃ-tʃʼag	42	낯짝	2	<i>unknown</i>
21	pit̚ʃʼulgi	pis-tʃul-ki	43	빗줄기	16	sheets of rain
		pit̚ʃʰ-tʃul-ki	44	빛줄기	2	rays of light
22	takt̚ʼal	takʼ-tal	47	닥달	8	scolding
		tak-tal	48	닥달	3	<i>unknown</i>
	15 actual sets (7 due to spelling errors); 32 nouns out of 34,803; 46,781 tokens out of 1,234,323					

- Each set possesses at least one Native Korean noun. This is to be expected, because Sino-Korean forms do not possess root-final release and/or laryngeal distinctions, and so are not subject to neutralizing aplosivization.
- The token frequency of one or more members of the homophonic sets is usually low: the mean token count among nouns is 35, but eight of the 15 homophonic sets possess at least one member with a token count below ten.
- While there are 46,781 homophonous tokens, a full 40,544 are of one word (것) (“one”), while its homophone (겉) (“surface”) has a token count of 120. If we exclude this one set—which, due to the high frequency of one word, and the low frequency of its homophone, is scarcely likely to induce confusion—then...
- ...Out of 1,234,323 noun tokens there are 6,117 homophonous tokens due to aplosivization.
- *This is a very low level of homophony.*

Interim summary:

- Korean aplosivization induces the neutralization of a remarkably high number of oppositions:** twelve values neutralize to three. 29% of the nouns are subject to neutralizing aplosivization.
- Korean aplosivization induces the homophony of a remarkably low number of nouns:** 15 sets of nouns are homophonous (32 nouns out of 34,803—<0.1%; 6,117 noun tokens out of 1,234,323 tokens—<0.1%).

(B) Nasal lateralization induces very little homophony

- n+r, l+r, l+n → l: (Kim-Renaud 1975, Martin 1992, Davis & Shin 1999). This process entered the language about 400 years ago (Martin 1992: 52).

(7) Number of nouns subject to nasal lateralization neutralization

Sequence	Number of derived sequences	Homophonic sets
lɾ	681 (from n+r)	10
	316 (from l+r)	
	4 (from l+n)	
Totals:	1001	10
1001 nasal-nasal sequences in the set of 34,803 nouns 10 homophonic sets		

(8) Homophonic sets due to nasal lateralization

Set number	Homophonous words	Non-neutralized allomorphs	Word number	Hangul	Token count	Gloss
1	tʃʰa <u>l</u> a	tʃʰa <u>l</u> -ra	1	찰라	5	<i>unknown</i>
		tʃʰa <u>l</u> -na	2	찰나	22	a moment
2	hwa <u>l</u> o	hwa <u>n</u> -ro	3	환로	2	file
		hwa <u>l</u> -ro	4	활로	11	bow
3	i <u>l</u> ju	i <u>n</u> -rju	5	인류	179	humanity
		i <u>l</u> -rju	6	일류	32	peculiarity
4	i <u>s</u> a <u>b</u> u <u>l</u> a <u>n</u>	i <u>s</u> -sa-pu <u>n</u> -ra <u>n</u>	7	일사분란	3	<i>unknown</i>
		i <u>s</u> -sa-pu <u>l</u> -ra <u>n</u>	8	일사불란	7	being in perfect order
5	ka <u>l</u> i <u>p</u> ʰ	ka <u>n</u> -ri <u>b</u>	9	건립	55	building
		ka <u>l</u> -ri <u>b</u>	10	걸립	9	alms rounds
6	kwa <u>l</u> i <u>j</u> a <u>n</u>	kw <u>a</u> n-rj <u>a</u> n	11	권련	2	deep affection
		kw <u>a</u> l-rj <u>a</u> n	12	궤련	10	cigarette
7	pu <u>l</u> i	pu <u>n</u> -ri	15	분리	124	separation
		pu <u>l</u> -ri	16	불리	7	handicap
8	pu <u>l</u> i <u>j</u> a <u>n</u>	pu <u>n</u> -rj <u>a</u> n	17	분량	61	quantity
		pu <u>l</u> -rj <u>a</u> n	18	불량	40	inferiority
9	sa <u>l</u> i <u>m</u>	sa <u>n</u> -ri <u>m</u>	19	산림	35	woodland
		sa <u>l</u> -ri <u>m</u>	20	살림	179	lifestyle
10	ʃi <u>l</u> o <u>k</u> ʰ	ʃi <u>n</u> -ro <u>g</u>	21	신록	3	fresh verdure
		ʃi <u>l</u> -ro <u>g</u>	22	실록	6	chronicle
11	su <u>l</u> e	su <u>l</u> -r <u>e</u>	23	술래	20	smell of liquor
		su <u>l</u> -n <u>e</u>	24	술내	4	tagger
12	ja <u>l</u> i <u>a</u> kʰ	ja <u>n</u> -ra <u>g</u>	25	연락	220	connection
		ja <u>l</u> -ra <u>g</u>	26	열락	2	joy
10 actual sets (2 due to spelling errors); 20 words; 1,001 tokens						

- There are ten homophonic sets. Two additional cases are false positives due to documented spelling errors.
- Eight of the actual sets possess at least one member with a token count under ten.
- Out of 1,234,323 tokens there are 1,001 homophonous tokens due to nasal lateralization
- *This is a very low level of homophony*

(C) Liquid nasalization induces very little homophony

- **p+r → mn, k+r → ɳn** (Kim-Renaud 1975, Martin 1992, Davis & Shin 1999).
- There is no phonetic motivation for the pattern. Rather, this alternation has the ‘feel’ of being analogically derived from nasal assimilation.

(9) Number of nouns subject to liquid nasalization neutralization

Sequence	Number of non-derived sequences	Number of derived sequences	Homophonic sets
m+n	74	45 (from p+r)	0
ɳ+n	394	182 (from k+r)	6
Totals:	468	227	6
695 nasal-nasal sequences in the set of 34,803 nouns			
6 homophonic sets			

- In all, there are six homophonic sets of nouns, and two which may be the result of spelling errors in the corpus. Excluding these two sets, a total of 520 nouns tokens (out of 1,234,323) are homophonous, which verges on 0% of the total.

(10) Homophonic sets due to liquid nasalization

Set number	Homophonous words	Non-neutralized allomorphs	Word number	Hangul	Token count	Gloss
1	tʃʌɳni	tʃʌɳ-ri	1	정리	287	arrangement
		tʃʌg-ri	2	적리	3	dysentery
2	tʃʌɳnipʰ	tʃʌɳ-rib	3	정립	71	triangular position
		tʃʌg-rib	4	적립	4	accumulation
3	kaɳnon	kaɳ-ron	5	강론	8	sermon
		kag-ron	6	각론	4	detailed exposition
4	kudʒiɳnan	ku-tʃig-nan	7	구직난	2	unemployment problem
		ku-tʃig-ran	8	구직란	2	unknown
5	kjʌɳni	kjʌɳ-ri	9	경리	14	unknown
		kjʌg-ri	10	격리	37	quarantine
6	peɳɳjan	peɳ-ɾjan	11	백련	15	white lotus
		peɳ-njan	12	백년	5	century

7	pjaŋŋjakʰ	pjaŋ-rjaŋ	13	병력	90	replacement depot
		pjaŋg-rjaŋ	14	벽력	4	thunder and lightning
8	jaŋŋjakʰ	jaŋ-rjaŋ	16	양력	26	solar calender
		jag-rjaŋ	16	약력	3	vita
6 actual sets (2 due to spelling errors); 12 words; 520 tokens						

- Out of 1,234,323 tokens there are 520 homophonous tokens due to liquid nasalization
- *This is a very low level of homophony*

(D) Nasal assimilation induces very little homophony

- Stops nasalize before nasals: p+N → mN, t+N → nN, k+N → ŋN

Word-internal nasal assimilation induces very little homophony

- Regarding word internal nasal sequences, 559 are derived consequence of nasal assimilation: 46 nouns possess m+m, 102 n+m, 271 ŋ+m, 18 m+n, 47 n+n, and 75 ŋ+n.
- There are 1,734 non-derived nasal+nasal sequences: 127 m+m, 473 n+m, 459 ŋ+m, 97 m+n, 184 n+n, and 394 ŋ+n. In all then, there are 2,293 nasal-nasal sequences in the set of 34,803 nouns.

(11) Number of nouns subject to nasal assimilation neutralization

sequence	Non-derived	Derived	Homophonic sets
m+m	127	46	0
n+m	437	102	2
ŋ+m	459	271	6
m+n	97	18	0
n+n	184	47	0
ŋ+n	394	75	2
Totals:	1,734	559	10
2,293 nasal-nasal sequences in the set of 34,803 nouns			
10 homophonic sets			

(12) Homophonic noun sets due to word-internal nasal assimilation

Set number	Homophonous words	Non-neutralized allomorphs	Word number	Hangul	Token count	Gloss
1	tjaŋmokʰ	tfaŋ-mog	1	작목	7	poor night vision
		tjaŋ-mog	2	장목	3	lumber
2	tjaŋmul	tfaŋ-mul	3	작물	35	crops
		tjaŋ-mul	4	장물	14	loot
3	tjaŋmun	tfaŋ-mun	5	작문	24	composition

		tʃaŋ-mun	6	장문	5	wide-open gate
4	tʃaŋnjan	tʃag-njan	7	작년	267	yesteryear
		tʃaŋ-njan	8	장년	8	prime of life
5	tʃʰanmul	tʃʰan-mul	9	찬물	26	cold water
		tʃʰas-mul	10	차물	2	tea
6	haŋmun	hag-mun	11	학문	234	scholarship
		haŋ-mun	12	항문	9	anus
7	koŋmul	kog-mul	13	곡물	20	corn
		koŋ-mul	14	공물	3	tribute
8	kudziŋnan*	ku-tʃig-nan	15	구직난	2	unemployment problem
		ku-tʃig-ran	16	구직란	2	unknown
9	kunmakʰ	kun-mag	17	군막	3	military tent
		kus-mag	18	굴막	3	miners' hut
10	orenman	o-reŋ-man	19	오랜만	116	unknown
		o-res-man	20	오랫만	8	unknown
11	peŋnjan	peg-rjan	21	백련	15	white lotus
		peg-njan	22	백년	5	century
12	paŋnamu	paŋ-na-mu	23	벚나무	5	cherry tree
		pas-na-mu	24	벗나무	2	unknown
13	saŋmul	saŋ-mul	25	성물	17	unknown
		sag-mul	26	석물	5	stone figures
14	jaŋmul	jag-mul	27	약물	42	medicinal waters
		jaŋ-mul	28	양물	7	penis
10 actual sets (4 due to spelling errors); 20 words; 732 tokens. *non-unique—already listed with lateral nasalization						

- In all, there are ten homophonic sets out of 2,293 neutralized nouns in the list of 34,803 nouns (732 out of 1,234,323 tokens), and four probable spelling errors.
- Out of 1,234,323 tokens there are 732 homophonous tokens due to word-internal nasal assimilation
- *This is a very low level of homophony*

Cross-word boundary nasal assimilation induces very little homophony

- Hwang (2008) investigates the amount of potential noun homophony due to the word boundary process.
- 51% of all nouns (17,763 out of 34,803) are potentially subject to neutralizing nasal assimilation (**P**=labial obstruent, **T**=coronal obstruent, **K**=dorsal obstruent).

(13) Potential word-final nasal-assimilated nouns

<u>Labials</u>			<u>Coronals</u>			<u>Dorsals</u>		
Value	Number of words	% of total	Value	Number of words	% of total	Value	Number of words	% of total
m#	1,913	5.3	n#	5,477	15.3	ŋ#	5,074	14.15
P#	1,343	3	T#	407	1	K#	3,549	10
Total:	3,256	8.7	Total:	5,884	16.4	Total:	8,623	23.9
17,763 out of 34,803 nouns; 51% of all nouns								

- Despite the enormous amount of potential neutralization here, Hwang shows that a mere 2.8% of the noun inventory is potentially subject to derived homophony as a consequence of nasal neutralization.

(14) Potential word-final nasal-assimilated homophones

<u>Labials</u>			<u>Coronals</u>			<u>Dorsals</u>		
Value	Number of words	% of total	Value	Number of words	% of total	Value	Number of words	% of total
m#	87	.2	n#	44	0.1	ŋ#	354	1.0
P#	89	.2	T#	59	0.1	K#	355	1.0
Total:	176	.4	Total:	103	0.25	Total:	709	2.0
988 words out of 34,803 nouns; 2.8% of all nouns								

- Any noun that might be subject to derived homophony as a consequence of nasal assimilation must be immediately followed by a nasal in the next word. This being the case, the amount of actual homophony is likely to be extremely low indeed.
- To get a sense of this value, there are 3,404 nouns in the corpus that are nasal-initial. Assuming this is representative of the language as a whole, this constitutes about 10% of the lexicon.
- This suggests that the likelihood of derived homophony is likely to be one-tenth of 2.8%, or .28%.
- This result is in full keeping with the remarkably low levels of derived homophony found elsewhere.

(E) Coronal assibilation induces very little homophony

- Coronal obstruents assibilate before **s** and **s'**, resulting in (neutralized) **s'** (Martin 1992).
- Out of the 34,803 nouns in the corpus, 131 words possess the relevant value: 41 lexical, 90 derived. There is a total of one homophonic set as a result of this alternation.

(F) Homophonic noun set due to coronal assibilation

Set number	Homophonous words	Non-neutralized allomorphs	Word number	Hangul	Token count	Gloss
1	pis'al	pitʃʰ-sal	1	빛살	12	light ray
		pis-sal	2	빗살	2	comb teeth
1 set; 2 words; 14 tokens						

- Out of 1,234,323 tokens there are 14 homophonous tokens due to coronal assibilation
- *This is a very low level of homophony*

(F) Cluster reinforcement induces no homophony

- When a non-aspirated obstruent comes to follow another obstruent, the second value tenses (Kim-Renaud 1975, Martin 1992).
- There are 4,048 nouns in the corpus that possess word-medial tensed obstruents: 449 **p'**, 476 **t'**, 1017 **tj'**, 1090 **s'**, and 1016 **k'**.
- All three sets of derived homophones are false positives.

(G) Homophonic noun sets due to cluster reinforcement

Set number	Homophonous words	Non-neutralized allomorphs	Word number	Hangul	Token count	Gloss
1	tʃ'ak <u>u</u> ŋ	tʃ'ag-k'uŋ	1	짜꿍	3	buddy
		tʃ'ag- <u>ku</u> ŋ	2	짜궁	2	<i>unknown</i>
2	k'akt' <u>u</u> gi	k'ag-tu-ki	3	깍두기	7	radish kimchi
		k'ag-t' <u>u</u> -ki	4	깍뚜기	4	<i>unknown</i>
3	kop' <u>e</u> gi	kob- <u>pɛ</u> -ki	5	곱배기	3	<i>unknown</i>
		kob- <u>p'</u> ɛ-ki	6	곱빼기	3	double shot
no actual set (all due to spelling errors); 0 words; 0 tokens						

- Out of 1,234,323 tokens there are no homophonous tokens due to cluster reinforcement
- *This is a very low level of homophony*

(G) Variable assimilation might induce more homophony (but might not...)

- In colloquial Korean speech, coronals (excluding the lateral) variably assimilate to a following consonant (Kim-Renaud 1975, Martin 1992), and labials variably assimilate to a following dorsal.
- This assimilation process has the potential to induce a non-negligible amount of homophony.

(16) Potential homophony due to variable assimilation

a. Variable value	Number of nouns	b. Non-variable value	Number of nouns	c. Number of potentially homophonic sets
t+P ~ p+P	126	p+P	92	No sets
n+P ~ m+P	683	m+P	385	15 sets
n+m ~ m+m	575	m+m	173	No sets
t+K ~ k+K	171	k+K	688	14 sets
p+K ~ k+K	177			
n+K ~ ŋ+K	1217	ŋ+K	1,324	62 sets
m+K ~ ŋ+K	249			
5,680 nouns				

91 potentially homophonic sets

- This is a *variable* pattern of neutralization: sometimes neutralization occurs (more often in casual speech), and sometimes it doesn't (more often in formal speech).
- Jun (1995): gestural reduction of labials (in **p+k** sequences) occurs about 35% of the time in casual speech, and about 15% of the time in formal speech. (See also Willis 2009)
- This variable pattern does not vary between discrete values **kɪ** versus **pk**. Rather the variation is gradient, such that tokens may, in theory, fall anywhere on the phonetic continuum between these two endpoints.
- This sort of variation sets up a situation in which *near-neutralizations* are practically inevitable: some speech tokens are nearly—but, crucially, not completely—neutralized, and so contrasts may remain recoverable despite the near-obliteration of their acoustic distinctiveness (for example, Dinnsen & Charles-Luce 1984, and Charles-Luce 1985 on Catalan, Port & O'Dell 1985, Charles-Luce 1985, Port & Crawford 1989 on German, Slowiaczek & Dinnsen 1985 on Polish, Pye 1986 on Russian, Warner, Jongman, Sereno, and Kems 2004 on Dutch, Bishop 2007 on Andalusian Spanish).
- Charles-Luce (1985): potential homophones in Catalan are more likely to remain acoustically distinct from each other in semantically ambiguous contexts, in comparison to both homophones in semantically transparent contexts, and to neutralized heterophones. Comparable findings are reported for Dutch (Warner, Jongman, Sereno, and Kems 2004).
- Anecdotally, an informal, non-systematic dictionary investigation of potentially homophonic sets due to variable coronal assimilation (approximately forty-five minutes checking an online dictionary 'Babylon Korean Dictionary', <http://www.babylon.com/dictionary/1271/Babylon-Korean-English.html>) yielded only a few false-positives for words with optionally assimilated sequences.
- Were these sequences genuinely neutralized, we might expect a higher number of false positives: completely neutralized forms are more likely subject to spelling errors
- The documented spelling errors in the Sejong Corpus are not random, but instead, always reflect accurate pronunciations of the (presumably intended) word.

(17) Summary of neutralizing alternations and derived homophony

Alternation	Number of nouns, both lexical and derived (out of 34,803)	Number of homophonic sets	Number of homophonic tokens (out of 1,234,323)
(A) Aplosivization	10,138	15	6,117 (46,781-40,664)
(B) Nasal lateralization	1,001	10	288
(C) Liquid nasalization	695	6	520
(D) Nasal assimilation	7,592	10	732
(E) Coronal assibilant	131	1	14

(F) Cluster reinforcement	4,048	0	0
(G) Variable assimilation	(5,680)	(91)	(undetermined)
Totals:	13,258	42	7,671

- In running Korean speech, out of 1,234,323 noun tokens encountered, chances are that about 7,671 will be homophonous.
- *This is a very low level of homophony*

“A very low level of homophony” compared to *what*?

- Totals such as those presented here acquire greater significance when they are compared to values that are *not* eligible for derived homophony (Surendran & Niyogi 2006).

True word-final aplosivization vs. would-be word-initial laryngeal neutralization

- Word-initial values provide a relevant comparison to word-final values, because the set of contrastive values allowed root-initially is very similar to (though larger than) the set of values allowed root-finally, and is, of course, not subject to aplosivization.

(18) *Would-be* homophonic sets due to word-initial laryngeal neutralization versus actual homophonic sets due to word-final aplosivization

<u>Labials</u>		<u>Coronals</u>		<u>Dorsals</u>	
(Would-be) Word-initial	(True) Word-final	(Would-be) Word-initial	(True) Word-final	(Would-be) Word-initial	(True) Word-final
4,323 words	1,202 words	15,988 words	408 words	5,814 words	3,524 words
100 sets	4 sets	1,755 sets	10 sets	57 sets	1 set
1,912 sets of would-be word-initial homophonic sets due to laryngeal neutralization vs. 15 true homophonic true sets due to aplosivization					

True stop+nasal assimilation vs. would-be nasal+stop assimilation

(19) Stop+nasal word-internal nasal assimilation versus *would-be* nasal+stop word-internal nasal assimilation

sequence	(Would-be) N+N sequences derived from nasal+stop sequences	(True) N+N sequences derived stop+nasal sequences	(True) Lexical N+N sequences
m+m	512	46	127
n+m	1179	102	473
ŋ+m	1236	271	459
m+n	911	18	97
n+n	3333	47	184
ŋ+n	4571	75	394
m+ŋ	249	--	--
n+ŋ	1238	--	--

η+η	1327	--	--
1100 would-be homophonic sets due to nasal-stop assimilation (1072 without +η) vs. 10 true sets due to stop-nasal assimilation			

Speculation: whence Korean neutralization?

- Korean has witnessed a massive influx of Chinese words in its distant past—beginning at least 1300 hundred years ago, with the greatest amount of borrowing during the Koryŏ Dynasty, about 1000 to 600 years ago—which served to supplant a significant portion of its native vocabulary, particularly its noun inventory.
- The augmentation of the Korean lexicon (due to contact with Chinese) induced phonotactic changes culminating in massive neutralization.
- Now, the phonotactic regularities seem to be inducing further lexical augmentation, setting up a “diachronic do-se-do” between the lexicon and the phonotactics.
- The simplification of root-final obstruent values continues into the present day:
- Whereas past neutralizing values have primarily been among *non-prevocalic* alternants, in present-day Korean there are additional neutralizations among *prevocalic* root-final values.
- Jun (2007): among the coronals, **s** is the favored variant; among the labials (**b**, **p^h**), **b** is the favored variant; among the dorsals (**g**, **k^h**, **k'**), **g** is the favored variant.
- These favored variants are also the most frequent in terms of the number of words that possess them,
- Jun: the *token prevalence* of some *variants* over others may be a consequence of the *lexical prevalence* of some *values* over others (See also Kim 2001, Ito 2006, Kang 2006, and Albright 2008 on this topic).
- The inventory of root-final values in Korean seems to be in a rather constant—if slow-going—state of attrition.
- The slow-going attrition of the Korean system of root-final values may have been set in motion by the influx of Sino-Korean forms.

(20) Diachrony of Korean root endings

(28) Summary of Korean Root Endings

Labials	Coronals						Dorsals	Commentary			
Korean consonantal root endings, lexically prevocalic, >600 years ago (bold values are also Sino-Korean):											
b	p^h	d	t^h	dz	ts^h	s	(h)	g	k^h	Chinese compounds enter	
Korean consonantal root endings, lexically non-prevocalic, >600 years ago:											
p^ʻ		t^ʻ		s		Ø		k^ʻ		Nasal assimilation begins	
Korean consonantal root endings, lexically prevocalic, >400 years ago:											
b	p^h	d	t^h	dz	ts^h	s	(h)	g	k^h	Nasal lateralization begins	
Korean consonantal root endings, lexically non-prevocalic, >400 years ago:											
p^ʻ		t^ʻ						k^ʻ			
Contemporary Korean consonantal root endings, lexically prevocalic (bold cells vary; bold values are preferred variants; tensed values arose from a collapse of “overstuffed” root endings):											
b	p^h	d	t^h	dʒ	tʃ^h	s	s^ʻ	(h)	g	k^h	k^ʻ
Contemporary Korean consonantal root endings, lexically non-prevocalic:											
p^ʻ		t^ʻ						k^ʻ		Root-final values merge	

- The attrition of root-final values was offset by the compounding nature of the Sino-Korean vocabulary. Two heads are better than one: compounding greatly increased the opportunity for nouns to contrast with each other.
- Proposal: the number of root-final values undergoing both neutralization and merger increased exactly because of the compounding that was introduced by the Sino-Korean vocabulary, which offset any counter-functional developments.

Final remarks

- The low level of homophony in Korean becomes especially obvious by comparing *actual* homophonic sets to *would-be* homophonic sets.
- Anti-homophony seems to be a genuine pressure acting on the Korean sound system.
- Even when there is a natural phonetic tendency toward a particular phonetic state, say non-prevocalic aplosivization among obstruents, a given language will most likely *not* evolve toward that state if excessive homophony were to result, for the very speech tokens that are produced with homophone-inducing aplosives are also the very speech tokens that would likely confuse listeners.
- As a natural, passive consequence, such tokens would not be reproduced as these listeners become speakers.
- There is likely to be a natural, passive maintenance of contrast—either through the maintenance of plosivization or by some other means—for it is exactly those speech tokens that are communicated successfully to listeners that are more likely to take hold, and become conventionalized as the speech norm.
- Successful speech variants, like successful mutations, are naturally selected.

It is not the desire to be understood, but rather the consequence of misunderstanding that influences language change. This mechanism implies a mismatch between producer and interpreter: the type of built-in instability that we would expect to find behind long-term shifts in language behavior. -William Labov (1994:586)

References

- Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Linguistic Universals and Language Change*. Oxford University Press.
- Bishop, Jason. 2007. Incomplete Neutralization in Eastern Andalusian Spanish: Perceptual Consequences of Durational Differences Involved in S-Aspiration. *Proceedings of the International Congress of Phonetic Sciences XVI*. 1765-1768.
- Bladon, Anthony. 1986. Phonetics for Hearers. In G. McGregor, ed., *Language for Hearers*. Oxford: Pergamon Press, 1-24.
- de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press
- Casenhiser, Devin M. 2005. Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language* 32, 319-343.
- Charles-Luce, Jan. 1985. Word-final devoicing in German: Effects of phonetic and sentential contexts. *Journal of Phonetics* 13, 309-324.
- Charles-Luce, Jan. 1993. The effects of semantic context on voicing neutralization. *Phonetica* 50, 28-43.
- Davis, Stuart, & Seung-Hoon Shin. 1999. The syllable contact constraint in Korean: an optimality-theoretic analysis. *Journal of East Asian Linguistics* 8, 285-312.
- Dinnsen, Daniel. 1985. A re-examination of phonological neutralization. *Journal of Linguistics*, 21, 265-279.
- Dinnsen, Daniel, & J. Charles-Luce, J. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12, 49-60.
- Cho Seung-Bog. 1967. *A Phonological Study of Korean*. Uppsala: Almqvist and Wiksells.
- Flemming, Edward. 1995. *Auditory Representations in Phonology*. Ph.D. dissertation, UCLA, published 2002, in Outstanding Dissertations in Linguistics series. New York: Routledge.
- Hockett, Charles F. 1955. *A Manual of Phonology*. Indiana University Publications in Anthropology and Linguistics 11.
- Hockett, Charles F. 1967. The quantification of functional load. *Word* 23, 320-339.
- van der Hulst, Harry & Jeroen van der Weijer. 1991. Topics in Turkish phonology. In H. Boeschoten & L. Verhoeven (eds.) *Turkish Linguistics Today*. Leiden: Brill. 11-59.
- Hwang, Jennifer. 2008. Neutralizing nasalization in Korean. unpublished manuscript, San José State University.
- Ito Chiayuki. 2006. Coronal Coda Distribution in Korean Nouns. unpublished manuscript, Tokyo University of Foreign Studies, ILCAA.
- Jun Jongho 1995 *A Constraint Based Analysis of Place Assimilation Typology*, Ph.D. Dissertation, UCLA.

- Jun Jongho. 2007. Variation in Korean stem-final obstruents. unpublished manuscript, SNU.
- Kang Yoonjung. 2006. Neutralizations and variations in Korean verbal paradigms. In Susumu Kuno et al. eds., *Harvard Studies in Korean Linguistics XI*.
- Kim Hyunsoon. 2001. A phonetically based account of phonological stop assimilation. *Phonology* 18, 81-108.
- Kim, Hyunsoon, & Allard Jongman. 1996. Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics* 24, 295-312.
- Kim-Renaud, Young-Key 1975. *Korean Consonantal Phonology*. Tower Press.
- King, Robert. 1967. Functional Load and Sound Change. *Language*, 43, 831-852.
- Kingston, John. 1985. *The Phonetics and Phonology of the Timing of Oral and Glottal Events*. Ph.D. dissertation, University of California at Berkeley.
- Kingston, John. 1990. Articulatory Binding. In John Kingston & Mary E. Beckman, eds., *Between the Grammar and Physics of Speech--Papers in Laboratory Phonology I*. Cambridge University Press, 406-434.
- Lee Iksop, & S. Robert Ramsey. 2000. *The Korean Language*. Cambridge: Cambridge University Press.
- Labov William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford, Blackwell.
- Lee June-Yub. 1994. *Hangul Code Conversion Program 2.1*. (<http://ftp.kaist.ac.kr/hangul/code/hcode/>).
- Lindblom, Bjorn, Peter MacNeilage, & Michael Studdert-Kennedy. 1984. Self-organizing processes and the explanation of language universals. In Brian Butterworth, Bernard Comrie, & Osten Dahl, eds., *Explanations for Language Universals*, Walter de Gruyter and Co., 181-203.
- Lombardi, Linda. 1991. *Laryngeal Features and Laryngeal Neutralization*, Ph.D. dissertation, U.Mass Amherst, published 1994, in Outstanding Dissertations in Linguistics series. New York: Garland.
- Martin, Samuel E. 1992. *A Reference Grammar of Korean*. Rutland, VT: Charles E. Tuttle Company.
- Martin, Samuel E. 1997. How did Korean get -l for Middle Chinese words ending in -t? *Journal of East Asian Linguistics* 6, 262-171.
- Martinet, Andre. 1952. Function, structure, and sound change. *Word* 8.2, 1-32.
- Ohala, John J. 1981. The listener as a source of sound change. In Carrie S. Masek, Roberta A. Hendrick, & Mary Frances Miller, eds., *Papers from the Parasession on Language and Behavior*, Chicago Linguistics Society, 178-203.
- Port, Robert, & Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17, 257-282.
- Port, Robert, & Michael L. O'Dell. 1985. Neutralization of syllable-final voicing in German *Journal of Phonetics* 13, 455-471.
- Pye, S. 1986. Word-final devoicing of obstruents in Russian. *Cambridge Papers in Phonetics and Experimental Linguistics* 5, 1-10.
- Silverman, Daniel. 1995. *Phasing and Recoverability*. Ph.D. dissertation, UCLA, published 1997, in Outstanding Dissertations in Linguistics series. New York: Garland.
- Silverman, Daniel. 1996. Phonology at the interface of morphology and phonetics: root-final laryngeals in Chong, Korean, and Sanskrit. *Journal of East Asian Linguistics* 5, 301-322.
- Silverman, Daniel. 2006. *A Critical Introduction to Phonology: of Sound, Mind, and Body*. London/New York: Continuum Books.
- Silverman, Daniel, & Jongho Jun. 1994. Aerodynamic evidence for articulatory overlap in Korean. *Phonetica* 51, 210-220.
- Slowiaczek, Louisa M., & Daniel Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13, 325-341.
- Sohn Ho-Minh. 1999. *The Korean Language*. Cambridge: Cambridge University Press.
- Song Jae Jung. 2005. *The Korean Language: structure, use and context*. New York: Routledge.
- Steriade, Donca. 1995. Positional Neutralization. unpublished manuscript, UCLA.
- Steriade, Donca. 1997 Phonetics in phonology: the case of laryngeal neutralization. unpublished manuscript, UCLA.
- Steriade, Donca. 2000 Directional asymmetries in place assimilation: a perceptual account. in Elizabeth Hume & Keith Johnson, eds., *Perception in Phonology*, Academic Press.
- Surendran Dinoj & Partha Niyogi. 2003. Measuring the usefulness (functional load) of phonological contrasts. *Technical Report TR-2003-12*, Department of Computer Science, University of Chicago.
- Surendran Dinoj, & Partha Niyogi. 2006. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In Ole Nødergaard Thomsen, ed., *Competing Models of Linguistic Change: Evolution and Beyond*. Amsterdam and Philadelphia: John Benjamins.
- Trubetskoy, Nikolai S. 1939 [1969]. *Principles of Phonology*. Berkeley/Los Angeles: University of California Press.
- Warner, Natasha, Allard Jongman, Joan Sereno, & Rachel Kemps. 2004. Incomplete neutralization and other sub-phonemic differences in production and perception: evidence from Dutch. *Journal of Phonetics* 32, 251-276.
- Wedel, Andrew. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23, 247-274.
- Wright, Richard. 2004. A Review of Perceptual Cues and Cue Robustness. In Bruce Hayes, Robert Kirchner, & Donca Steriade (eds.), *Phonetically Based Phonology*. Cambridge, UK: Cambridge University Press.