*J. Linguistics* **46** (2010), 453–482. © Cambridge University Press 2009
doi:10.1017/S0022226709990247 First published online 20 November 2009

# Neutralization and anti-homophony in Korean[1]

DANIEL SILVERMAN

*San José State University*

(Received 18 August 2008; revised 27 June 2009)

Neutralization in Korean involves a large number of oppositions, and affects a significant portion of the lexicon. Nonetheless, it induces remarkably little homophony. These highly divergent facts are argued herein to be related. 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)

Generally speaking, low-level phonetic variation among contrastive values is passively delimited such that a comfortable perceptual 'buffer zone' is maintained between one value and its immediate systemic 'neighbors' (see e.g. Martinet 1952; Hockett 1955, 1967; Lindblom, MacNeilage & Studdert-Kennedy 1984; Labov 1994; Flemming 1995; de Boer 2001; Silverman 2006). However, under certain circumstances this buffer zone can be fatally breached such that one value may (diachronically) merge or (synchronically) neutralize with another. Martinet (1952), for example, observes that there is a strong tendency for merged values to have been phonological neighbors in the past, in the sense that they had already shared a significant number of phonetic features. For Martinet, though, the fact that a language possesses similar values does not necessarily increase the likelihood of these values' merging or neutralizing. What crucially matters, according to Martinet, is the extent to which this opposition minimally distinguishes meaningful units of the language (minimal pairs), since the amount of homophony in a lexicon may have an impact on communicative success: the more homophony, the more likely that the speech signal will contain lexical ambiguities, and so the more

[1] For their help (which does not entail their endorsement of my conclusions), thanks to Adam Albright, Devin Casenhiser, Stuart Davis, Ik-Sang Eom, David Etienne-Bouchard, Andrew Garrett, Chris Golston, Bill Idsardi, Chiyuki Ito, Jongho Jun, Yoonjung Kang, Abby Kaplan, Bob Kennedy, Hahn Koo, Bruce Lyon, Öner Özçelik, Glyne Piggott, Richard Wiese, members of my SJSU Phonology 2 class S08, and two anonymous *Journal of Linguistics* referees. This paper is dedicated to the memory of my teacher, Peter Ladefoged.

often speech might induce confusion in listeners. Under the assumption that successful speech is more likely to establish itself as the conventionalized norm than is unsuccessful speech, then *ceteris paribus* speech with fewer homophones is more likely to establish itself as the conventionalized norm than is speech with more homophones. In all, according to Martinet, the tendency toward merger or neutralization 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 (in the sense of Trubetzkoy 1939/1969).
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.

Ultimately, however, merger or neutralization is not likely to take place simply because some of the relevant morphemes have a low token frequency, or because some of the relevant minimal pairs happen to belong to different syntactic categories, etc. Rather if sounds are to merge or neutralize, then, according to Martinet's 'functional load' criteria, the output of the process – across the lexicon in use – should not yield a significant increase in the amount of communicative confusion.

Armed with these intuitively appealing criteria for the determination of functional load, Martinet is nonetheless quick to retreat:

> It is clear that the functional yield of an opposition can only be evaluated with any degree of accuracy if we deal with linguistic stages for which fairly exhaustive word lists are available. This circumstance makes it practically impossible to check the validity of the functional assumption in the case of prehistoric sound shifts. (Martinet 1952: 9)

It is not the purpose of the present study to investigate the role that 'functional load' may or may not play in (diachronic) sound mergers; Martinet's concerns appear fully justified in that, even at our present level of knowledge, extensive information about the lexicon in use is typically exceedingly difficult to ascertain for two different stages of a language, i.e. both before and after a merger takes place (but see Surendran & Niyogi 2003, 2006 for an example from Cantonese; see also King 1967 and Hockett 1967, who disagree with each other on the merits of Martinet's proposals). Instead, in this paper I report on a case study of (synchronic) neutralization and its functional consequences. With today's computer-tallied corpora, such an investigation can be readily undertaken.

Taking all of Martinet's criteria into account, the hypothesis to be explored herein is straightforward:

> Neutralizing alternations are more likely to be present in a language if their outputs do not significantly increase the level of homophony; derived homophony should not be 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. While this value may ultimately be ascertained by thoroughly investigating many languages, I intend merely to show that, in a given language, the amount of derived homophony is remarkably low.

The overall goal of the present study is to reintroduce Martinet's proposals into modern discourse on phonological theory, and to provide a case study showing that a language may indeed possess significant amounts of neutralization with negligible counter-functional consequences. The focus of this paper is neutralization in Korean. Indeed, Korean is the prototypical 'Linguistics 101' example of neutralization, as neutralization here involves so many contrastive values.

Section 1 investigates the extent of potential homophony in Korean due to six neutralizing alternations: aplosivization, nasal lateralization, liquid nasalization, cluster reinforcement, assibilation, and nasal assimilation (Martin 1992), and shows that these induce very little homophony. Also briefly considered is variable assimilation of labials-to-dorsals, and coronals-to-labials and coronals-to-dorsals. Overall, it is found that, despite extensive neutralization, derived homophony in Korean is remarkably rare.

In section 2, actual patterns of neutralization in Korean are compared to hypothetical neutralizing patterns, including would-be word-initial laryngeal neutralization, would-be nasal lateralization, and would-be nasal + stop assimilation. It is shown that these hypothetical patterns, had they evolved in Korean, would have induced far more homophony than is actually found.

In section 3, some of the historical pressures that seem to have influenced the present-day Korean pattern are briefly considered, especially the role that Chinese has played in the history of the Korean sound pattern; I further discuss current developments in Korean, suggesting that these historical trends are continuing into the present. I speculate that 'functional load' (in the sense of Martinet) might indeed play a role in the diachronic comportment of linguistic sound systems.

## 1. Neutralization in Korean

Table 1 provides intervocalic consonantal values in Korean (adapted from Martin 1992). The first column lists morpheme-final consonants; the first row lists following morpheme-initial consonants, and the table's interior

| +C→ C+↓ | p | pʰ | p’ | t | tʰ | t’ | s | s’ | ʧ | ʧʰ | ʧ’ | k | kʰ | k’ | m | n | ɾ | h | ∅ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| p | p’ | pʰ | p’ | pt’ | ptʰ | pt’ | ps’ | ps’ | pʧ’ | pʧʰ | pʧ’ | k’ | kʰ | k’ | m: | mn | mn | pʰ | b |
| pʰ | p’ | pʰ | p’ | pt’ | ptʰ | pt’ | ps’ | ps’ | pʧ’ | pʧʰ | pʧ’ | k’ | kʰ | k’ | m: | mn | mn | pʰ | pʰ |
| t | p’ | pʰ | p’ | t’ | tʰ | t’ | s’ | s’ | ʧ’ | ʧʰ | ʧ’ | k’ | kʰ | k’ | m: | l: | l: | tʰ | d |
| tʰ | p’ | pʰ | p’ | t’ | tʰ | t’ | s’ | s’ | ʧ’ | ʧʰ | ʧ’ | k’ | kʰ | k’ | m: | n: | n: | tʰ | tʰ |
| s | p’ | pʰ | p’ | t’ | tʰ | t’ | s’ | s’ | ʧ’ | ʧʰ | ʧ’ | k’ | kʰ | k’ | m: | n: | n: | s’ | s |
| s’ | p’ | pʰ | p’ | t’ | tʰ | t’ | s’ | s’ | ʧ’ | ʧʰ | ʧ’ | k’ | kʰ | k’ | m: | n: | n: | s’ | ʃ’ |
| ts | p’ | pʰ | p’ | t’ | tʰ | t’ | s’ | s’ | ʧ’ | ʧʰ | ʧ’ | k’ | kʰ | k’ | m: | n: | n: | ʧʰ | ʤ |
| ʧʰ | p’ | pʰ | p’ | t’ | tʰ | t’ | s’ | s’ | ʧ’ | ʧʰ | ʧ’ | k’ | kʰ | k’ | m: | n: | n: | tʰ | ʧʰ |
| k | kp’ | kpʰ | kp’ | kt’ | ktʰ | kt’ | ks’ | ks’ | kʧ’ | kʧʰ | kʧ’ | k’ | kʰ | k’ | ŋm | ŋn | ŋn | kʰ | g |
| kʰ | kp’ | kpʰ | kpʰ | kt’ | ktʰ | kt’ | ks’ | ks’ | kʧ’ | kʧʰ | kʧ’ | k’ | kʰ | k’ | ŋm | ŋn | ŋn | kʰ | kʰ |
| k’ | kp’ | kpʰ | kp’ | kt’ | ktʰ | kt’ | ks’ | ks’ | kʧ’ | kʧʰ | kʧ’ | k’ | kʰ | k’ | ŋm | ŋn | ŋn | kʰ | k’ |
| m | mb | mpʰ | mb | md | mtʰ | mt’ | ms’ | ms’ | mʧ’ | mʧʰ | mʧ’ | ŋk’ | ŋkʰ | ŋk’ | m: | mn | mn | mh | m |
| n | mb | mpʰ | mp’ | nd | ntʰ | nt’ | ns’ | ns’ | nʧ’ | nʧʰ | nʧ’ | ŋk’ | ŋkʰ | ŋk’ | m: | n: | n: | nh | n |
| ŋ | ŋb | ŋpʰ | ŋp’ | ŋg | ŋtʰ | ŋt’ | ŋs’ | ŋs’ | ŋʧ’ | ŋʧʰ | ŋʧ’ | ŋk’ | ŋkʰ | ŋk’ | ŋm | ŋn | ŋl | ŋh | ŋ |
| l | lb | lpʰ | lp’ | ld | ltʰ | lt’ | ls’ | ls’ | lʧ’ | lʧʰ | lʧ’ | lg | lkʰ | lk’ | lm | l: | l: | lh | l |
| ∅ | b | pʰ | p’ | d | tʰ | t’ | s’ | s’ | ʤ | ʧʰ | ʧ’ | g | kʰ | k’ | m | n | ɾ | ∅ | ∅ |

*Table 1*
Korean intervocalic values (shaded cells are variably implemented)

indicates the phonetic values of all combinations of these morphologically-ordered elements.

In all, there are 304 morphologically-ordered sequences here. Variable values are in shaded boxes. Including these variable values among the neutralized values for now, the 304 morphologically ordered sequences reduce to only 75 phonetic values, a reduction of about 75%. In this section I show that, despite this drastic reduction in the number of intervocalic values, derived homophony remains minimal. I consider in turn (1.1) aplosivization, (1.2) nasal lateralization, (1.3) liquid nasalization, (1.4) nasal assimilation, (1.5) coronal assibilation, (1.6) cluster reinforcement, and (1.7) variable assimilation.

The data for this study were tallied from the Sejong Corpus, a database collected from written sources (http://sejong.or.kr/). The version of the Sejong Corpus used here is the same as that used in Albright (2008). The Hangul was Romanized by Albright using HCode Hangul Code Conversion software (Lee 1994). As noun roots often appear unsuffixed as well as suffixed in Korean (unlike verb roots, which are obligatorily suffixed), they are far more likely to engage in neutralizing alternations involving their final consonantal elements. For this reason, only nouns are considered herein.

All told, out of 35,907 distinct nouns in the corpus, the six categorical neutralizing alternations (below) are shown to derive only 42 sets of homophones, involving 86 words. This is a very low level of homophony.

### 1.1 *Aplosivization induces very little homophony*

Cross-linguistically, 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). If a stop is not released into a more open gesture such as a vowel, it may lose the phonetic cues associated with this interval of the speech stream, among them, cues to the state of the larynx. In the limiting case, the perceptual distinction among contrastive laryngeal states is extinguished completely. This is laryngeal neutralization due to aplosivization.

Table 2 displays the pattern of aplosivization in Korean, a pattern that seems to have gradually entered Korean during and after The Kolye Dynasty, from about 1000 to 600 years ago, the era of extensive borrowing from Chinese (Martin 1992). (Shaded values do not alternate, as they never appear in morpheme-final position.) The plain obstruents, the aspirated plosives (and **h**), and the tensed obstruents may appear before vowels in Korean. Aplosives are found elsewhere, that is, in lexical non-prevocalic contexts. Due to the rich suffixing system in Korean, the plosives (except for

| | Lexically prevocalic | | | | Lexically non-prevocalic |
|---|---|---|---|---|---|
| | Plain | Aspirated | Tensed | | Aplosive |
| Labial | **p** | **p$^{h}$** | **p'** | Alternates with | **p˺** |
| Coronal | **t** | **t$^{h}$** | **t'** | | **t˺** |
| | **ʧ** | **ʧ$^{h}$** | **ʧ'** | | |
| | **s** | **(h)** | **s'** | | |
| Dorsal | **k** | **k$^{h}$** | **k'** | | **k˺** |

*Table 2*
Distribution of Korean obstruents (and **h**)

| | Plosive | | Alternates with | Aplosive | |
|---|---|---|---|---|---|
| | (a) Locative (-e) | (b) Nominative (-i) | | (c) Isolation Form | Gloss |
| Labial | **pa<u>b</u>-e** | **pa<u>b</u>-i** | | **pa<u>p</u>˺** | rice |
| | **i<u>p</u>$^{h}$-e** | **i<u>p</u>$^{h}$-i** | | **i<u>p</u>˺** | leaf |
| Coronal | **o<u>s</u>-e** | **o<u>ʃ</u>-i** | | **o<u>t</u>˺** | clothes |
| | **pa<u>t</u>$^{h}$-e** | **pa<u>ʧ</u>$^{h}$-i** | | **pa<u>t</u>˺** | field |
| | **na<u>dʒ</u>-e** | **na<u>dʒ</u>-i** | | **na<u>t</u>˺** | day |
| | **i<u>tʃ</u>$^{h}$-e** | **i<u>tʃ</u>$^{h}$-i** | | **i<u>t</u>˺** | light |
| Dorsal | **ku<u>g</u>-e** | **ku<u>g</u>-i** | | **ku<u>k</u>˺** | soup |
| | **puə<u>k</u>$^{h}$-e** | **puə<u>k</u>$^{h}$-i** | | **puə<u>k</u>˺** | kitchen |
| | **pa<u>k</u>'-e** | **pa<u>k</u>'-i** | | **pa<u>k</u>˺** | outside |

*Table 3*
Examples of plosive-aplosive alternation in Korean

the tense anterior plosives), the fricatives, as well as **h**, all alternate with the aplosives. Two values neutralize to the labial aplosive, seven values neutralize to the coronal aplosive, and three values neutralize to the dorsal aplosive. Altogether, twelve values neutralize to three, which constitutes a remarkably high 75% reduction in the number of contrasts here.

Some examples of alternating values are provided in table 3 (adapted from Jun 2007). In table 3, note especially that there is an array of laryngeal distinctions in root-final position when a vowel-initial suffix is added (3a, b;

| Labials | | Coronals | | Dorsals | |
|---|---|---|---|---|---|
| Non-neutralized alternant | Number of words | Non-neutralized alternant | Number of words | Non-neutralized alternant | Number of words |
| **p**# | 1,154 | **t**# | 3 | **k**# | 3,522 |
| **p**+ | 762 | **t**+ | 0 | **k**+ | 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 |
| | | **ʤ**# | 10 | | |
| | | **ʤ**+ | 41 | | |
| | | **ʧ$^{h}$**# | 74 | | |
| | | **ʧ$^{h}$**+ | 43 | | |
| | | **ʧ**'# | 0 | | |
| | | **ʧ**'+ | 0 | | |
| | | **s**# | 257 | | |
| | | **s**+ | 612 | | |
| | | **(h)**# | 0 | | |
| | | **(h)**+ | 5 | | |
| | | **s**'# | 0 | | |
| | | **s**'+ | 2 | | |
| Totals | | | | | |
| **p̚**# | 1,343 | **t̚**# | 407 | **k̚**# | 3,549 |
| **p̚**+ | 795 | **t̚**+ | 763 | **k̚**+ | 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 35,907 nouns; 28% of all nouns | | | | | |

*Table 4*
Distribution of word-final (#) and morpheme-final (+) obstruents/h for 10,138 nouns, from the Sejong Project

underlined). However, these distinctions all neutralize to their corresponding aplosive value when lexically non-prevocalic (3c; also underlined). Note that coronals palatalize before **i**, and plain plosives voice between sonorants.

All told, the Sejong Project corpus lists 35,907 different nouns (not the number of tokens). Of this total, 10,138 nouns (28% of the total) possess aplosive alternants: 5,299 are word-final, and 4,839 are word-internal. These are pooled according to their word- and morpheme-final plosive (i.e. non-neutralized) alternant in table 4 (these would be 'underlying forms' in a

| | | | |
|---|---|---|---|
| **nɑd͡ʒ-i** | day (Nominative) | **nɑt̚-k'wɑ** | day and |
| **nɑt͡ʃʰ-i** | face (Nominative) | **nɑt̚-k'wɑ** | face and |
| **nɑʃ-i** | sickle (Nominative) | **nɑt̚-k'wɑ** | sickle and |

*Table 5*
Derived homophony in Korean

theory that hypothesizes the existence of this abstract level of representation). In principle, a given form can be counted more than once if, for example, it possesses both word-internal and word-final aplosive alternants, though this rarely seems to occur. At 28%, it is clear that the proportion of the noun vocabulary that is subject to neutralizing aplosivization is not insignificant.

Korean indeed possesses nouns that are rendered homophonous as a consequence of neutralizing aplosivization. An oft-cited example of derived noun homophony is provided in table 5. The Native (i.e. non-Sino-Korean) Korean words for 'day,' 'face', and 'sickle' are phonetically distinct when suffixed with the vowel-initial suffix **i** (the Nominative marker). However, when any of these words is suffixed with a consonant-initial suffix such as **kwɑ** 'and', the root-final obstruent aplosivizes to **t̚**, and the words are rendered phonetically non-distinct.

Among the 10,138 nouns that are subject to aplosivization, the sets in table 6 comprise an exhaustive list of the homophonous words. Certain words in the corpus have eluded a dictionary search, and are unfamiliar to native speakers of Korean, and these are marked 'unknown' – these sets are shaded. These facts, coupled with the very low token frequencies here, suggest that these are mere spelling errors documented in the Sejong Project corpus. (The significance of these errors is discussed in section 1.7.) Not including these suspect forms, in total, only 15 sets (32 words in all) – out of a total of 35,907 nouns in the corpus – are subject to derived homophony due to aplosivization.

It should be emphasized that the Sejong Project corpus tallies fully inflected words across the lexicon in use, and not simply bare roots. Consequently, it is not the case that all possible derived homophonic roots are documented here. Indeed, certain known potentially homophonous roots are absent from the corpus in their bare form, for example, **otʃʰ** 'lacquer' – **os** 'clothing'. Rather, the list in table 6 consists of those words documented in the Sejong Project corpus – whether bare roots or inflected – that are homophonous as a consequence of neutralizing aplosivization. This is an important point, because the present investigation intends to document homophony across the lexicon in use, and not across the lexicon *in toto*. The exhaustive list in table 6 makes it clear that the counter-functional

consequences of neutralizing aplosivization among nouns must be exceedingly meager: only 32 out of 35,907 nouns are homophonous due to aplosivization.

Moreover, words that engage in aplosivization typically have a low token count. In table 6, I provide the token counts of all the relevant forms, tallied from the Sejong Project corpus. While the mean token count among nouns is 35, ten of the 15 noun sets possess at least one member with a token count below ten. While there are 46,781 tokens, a full 40,544 are of one word (것), while its homophone (겉) has a token count of 120. If we discount this one set – which, due to the high frequency of the one word and the low frequency of its homophone, is scarcely likely to induce confusion – then, out of 1,234,323 noun tokens, the total number of homophonic tokens due to aplosivization reduces to 6,117.

To summarize, the most salient – and saliently divergent – findings of our discussion are:

(i) Korean aplosivization induces the neutralization of a remarkably high number of oppositions: twelve values neutralize to three; 28% of the nouns are subject to neutralizing aplosivization, and yet (ii) Korean aplosivization induces the homophony of a remarkably low number of nouns: 15 sets of nouns are homophonous (32 nouns out of 35,907: <0.1%; 6,117 noun tokens out of 1,234,323 tokens: <0.1%). This is a very low level of homophony.

### 1.2 *Nasal lateralization induces very little homophony*

Coronal nasals and laterals bear certain perceptual similarities, since their formant transitions are rather comparable, and they both possess formants and anti-formants. In Korean, their perceptual confusability has apparently culminated in a neutralizing alternation: a sequence of a lateral or tap and a coronal nasal in either order is realized as a long lateral: **n**+ɾ, **l**+ɾ, **l**+**n**→**l:** (Kim-Renaud 1975, Martin 1992, Davis & Shin 1999). This process entered the language about 400 years ago (Martin 1992: 52).

As the non-prevocalic lateral alternates with the tap when prevocalic, the acoustic connection between this latter alternant and the coronal nasal becomes rather tenuous. Though aerodynamically a non-obstruent (since the duration of its oral closure is insufficient to significantly reduce transglottal airflow), the tap possesses neither formants nor anti-formants; it is an extremely short oral stop. However, critical to the implementation of a tap is an oral opening both after and before the oral closure: a tap is only a tap if the tongue makes BRIEF contact with a passive articulator. In the context of a preceding gesture that involves prolonged contact between these two articulators – such as a coronal nasal or a lateral – there is little opportunity for the tongue to fall away from its contact site, and then quickly re-implement contact. Rather, tongue contact is likely to be

| Set number | Homophonous value | Non-neutralized allomorphs | Word number | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 1 | tʃʌt̚ | tʃʌtʃ | 1 | 젖 | 44 | breast/milk |
| | | tʃʌs | 2 | 젓 | 5 | salted fish |
| 2 | tʃʌtkaɾak̚ | tʃʌs-ka-ɾak | 3 | 젓가락 | 27 | chopsticks |
| | | tʃʌtʃ-ka-ɾak | 4 | 젖가락 | 3 | unknown |
| 3 | tʃip̚ | tʃip | 5 | 집 | 9 | house |
| | | tʃiph | 6 | 짚 | 19 | straw |
| 4 | tʃot̚ | tʃotʃ | 7 | 좇 | 21 | penis |
| | | tʃotʃh | 8 | 좇 | 2 | unknown |
| 5 | ip̚ | ip | 9 | 입 | 1,139 | mouth |
| | | iph | 10 | 잎 | 158 | leaf |
| 6 | kʌt̚ | kʌs | 11 | 것 | 40,544 | one |
| | | kʌth | 12 | 겉 | 120 | surface |
| 7 | mat̚ | mas | 13 | 맛 | 392 | taste |
| | | math | 14 | 말 | 3 | nearby place |
| 8 | miltʃ'ip̚ | mil-tʃip | 15 | 밀집 | 21 | crowd |
| | | mil-tʃiph | 16 | 밀짚 | 2 | straw |
| 9 | nat̚ | natʃ | 17 | 낮 | 294 | day |
| | | natʃh | 18 | 낯 | 91 | face |
| | | nas | 19 | 낫 | 25 | sickle |
| 10 | odʒiɾap̚ | o-tʃi-ɾap | 20 | 오지랍 | 4 | unknown |
| | | o-tʃi-ɾaph | 21 | 오지랖 | 3 | front of outer garment |
| 11 | pak̚ | pak | 22 | 박 | 581 | gourd |
| | | pak' | 23 | 밖 | 1,568 | exterior |
| 12 | pat̚ | path | 24 | 밭 | 380 | heritage |
| | | pat | 25 | 받 | 2 | field |
| 13 | pit̚ | pitʃh | 26 | 빛 | 726 | light |
| | | pitʃ | 27 | 빚 | 128 | debt |
| | | pis | 28 | 빗 | 6 | comb |

*Table 6*
*(Continued)*

| Set number | Hom-ophonous value | Non-neutralized allomorphs | Word number | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 14 | pʲʌt˺ | pʲʌtʰ | 29 | 볕 | 32 | sunshine |
| | | pʲʌs | 30 | 볏 | 6 | crest (of fowl) |
| 15 | sɛudʒʌt˺ | sɛ-u-ʧʌs | 31 | 새우젓 | 8 | unknown |
| | | sɛ-u-ʧʌʧ | 32 | 새우젖 | 2 | salted shrimp |
| 16 | sut˺ | suʧʰ | 33 | 숯 | 14 | charcoal |
| | | sutʰ | 34 | 숱 | 7 | hair density |
| 17 | tʌt˺ | tʌʧʰ | 35 | 덫 | 31 | trap |
| | | tʌs | 36 | 덧 | 4 | a short time |
| 18 | tʰe-ip˺ | tʰe-ipʰ | 37 | 테잎 | 2 | tape |
| | | tʰe-ip | 38 | 테입 | 9 | unknown |
| 19 | ʧipt'an | ʧip-tan | 39 | 집단 | 400 | group |
| | | ʧipʰ-tan | 40 | 짚단 | 12 | sheaf of straw |
| 20 | naʧ'ak˺ | naʧʰ-ʧ'ak | 41 | 낯짝 | 5 | face |
| | | naʧ-ʧ'ak | 42 | 낯짝 | 2 | unknown |
| 21 | pitʃ'ulgi | pis-ʧul-ki | 43 | 빗줄기 | 16 | sheets of rain |
| | | piʧʰ-ʧul-ki | 44 | 빛줄기 | 2 | rays of light |
| 22 | takt'al | tak'-tal | 45 | 닦달 | 8 | scolding |
| | | tak-tal | 46 | 닥달 | 3 | unknown |
| 15 actual sets (7 due to spelling errors); 32 nouns; 46,781 tokens | | | | | | |

*Table 6*
Homophonic noun sets due to aplosivization (shaded cells possess spelling errors)

maintained – much like aplosivization – resulting in a single gesture. This results in either a longer lateral or a longer coronal nasal. Nasal lateralization, consequently, may be seen as an analogous to aplosivization: due to the absence of lateral or nasal release, tongue contact is maintained for the duration of the morphologically-ordered consonantal sequence. Since the lateral and the coronal nasal are confusable, the origins of this neutralization pattern now come into focus.

Table 7 tallies the number of derived **l:** sequences found in the corpus. Out of 1,001 nouns in the corpus possessing **l:**, 681 derive from **n**+**ɾ**, 316 derive

| Sequence | Number of derived sequences | Homophonic sets |
|---|---|---|
| **lː** | 681 (from **n**+**ɾ**) | 10 |
| | 316 (from **l**+**ɾ**) | |
| | 4 (from **l**+**n**) | |
| 1001 sequences in the set of 35,907 nouns; 10 homophonic sets | | |

*Table 7*
Derived **lː** sequences

from **l**+**ɾ** and four derive from **l**+**n**. Table 8 provides an exhaustive list of derived homophonic sets due to nasal lateralization. There are ten homophonic sets, and 1001 homophonic tokens. Two additional cases are probably spelling errors. Seven of the actual sets possess at least one member with a token count under ten.

### 1.3 *Liquid nasalization induces very little homophony*

Another neutralizing alternation in Korean involves the nasalization of non-coronal obstruent – liquid (tap) sequences: **p**+**ɾ**→**mn**, **k**+**ɾ**→**ŋn** (Kim-Renaud 1975, Martin 1992, Davis & Shin 1999). This alternation is mentioned only in passing by Martin (1992); he does not discuss when the pattern may have entered the language. Surely there is no phonetic motivation for the pattern. Rather, this alternation has the 'feel' of being analogically derived from nasal assimilation (to be discussed in section 1.4). It is likely, therefore, that the alternation is a relatively recent innovation.

In all, 695 nouns listed in the corpus possess the relevant sequences. As shown in table 9, there are 119 nouns possessing **mn** (45 derive from **p**+**ɾ** and 74 are lexical **m**+**n**) and 576 nouns possessing **ŋn** (182 derive from **k**+**ɾ** and 394 are lexical **ŋ**+**n**).

In table 10 I provide an exhaustive list of homophones due to liquid nasalization. There are eight cases in all, two of which are suspected spelling errors. Excluding these two sets, a total of 520 nouns tokens (out of 1,234,323) are homophonous, which verges on 0% of the total.

### 1.4 *Nasal assimilation induces very little homophony*

Korean has a process of nasal assimilation such that any obstruent that comes to precede a nasal becomes nasal itself (Kim-Renaud 1975, Martin 1992,

| Set number | Hom-ophonous value | Non-neutralized allomorphs | Word number | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 1 | **ʧʰal:a** | **ʧʰal-ɾa** | 1 | 찰라 | 5 | unknown |
| | | **ʧʰal-na** | 2 | 찰나 | 22 | a moment |
| 2 | **hwal:o** | **hwan-ɾo** | 3 | 환로 | 2 | file |
| | | **hwal-ɾo** | 4 | 활로 | 11 | bow |
| 3 | **il:ʲu** | **in-ɾʲu** | 5 | 인류 | 179 | humanity |
| | | **il-ɾʲu** | 6 | 일류 | 32 | peculiarity |
| 4 | **ilsabul:an** | **il-sa-pun-ɾan** | 7 | 일사분란 | 3 | unknown |
| | | **il-sa-pul-ɾan** | 8 | 일사불란 | 7 | being in perfect order |
| 5 | **kʌl:ip˺** | **kʌn-ɾip** | 9 | 건립 | 55 | building |
| | | **kʌl-ɾip** | 10 | 걸립 | 9 | alms rounds |
| 6 | **kwʌl:ʲʌn** | **kwʌn-ɾʲʌn** | 11 | 권련 | 2 | deep affection |
| | | **kwʌl-ɾʲʌn** | 12 | 궐련 | 10 | cigarette |
| 7 | **pul:i** | **pun-ɾi** | 13 | 분리 | 124 | separation |
| | | **pul-ɾi** | 14 | 불리 | 7 | handicap |
| 8 | **pul:ʲaŋ** | **pun-ɾʲaŋ** | 15 | 분량 | 61 | quantity |
| | | **pul-ɾʲaŋ** | 16 | 불량 | 40 | inferiority |
| 9 | **sal:im** | **san-ɾim** | 17 | 산림 | 35 | woodland |
| | | **sal-ɾim** | 18 | 살림 | 179 | lifestyle |
| 10 | **ʃil:ok˺** | **ʃin-ɾok** | 19 | 신록 | 3 | fresh verdure |
| | | **ʃil-ɾok** | 20 | 실록 | 6 | chronicle |
| 11 | **sul:ɛ** | **sul-ɾɛ** | 21 | 술래 | 20 | smell of liquor |
| | | **sul-nɛ** | 22 | 술내 | 4 | tagger |
| 12 | **jʌl:ak˺** | **jʌn-ɾak** | 23 | 연락 | 220 | connection |
| | | **jʌl-ɾak** | 24 | 열락 | 2 | joy |
| 10 actual sets (2 due to spelling errors); 20 words; 1,001 tokens | | | | | | |

*Table 8*
Homophonic noun sets due to nasal lateralization

| Sequence | Number of non-derived sequences | Number of derived sequences | Homophonic sets |
|---|---|---|---|
| **m** + **n** | 97 | 45 (from **p** + **ɾ**) | 0 |
| **ŋ** + **n** | 394 | 182 (from **k** + **ɾ**) | 6 |
| Totals | 468 | 227 | 6 |
| 695 nasal + nasal sequences in the set of 35,907 nouns 6 homophonic sets | | | |

*Table 9*
Number of nouns subject to liquid nasalization neutralization

Davis & Shin 1999). This process entered Korean at least 600 years ago (Martin 1992: 52). There are two domains of application of this pervasive process: across morpheme boundaries, and across word boundaries (Kim-Renaud 1975).

Regarding word-internal nasal sequences, 559 are derived through nasal assimilation: 46 nouns possess **m** + **m**, 102 **n** + **m**, 271 **ŋ** + **m**, 18 **m** + **n**, 47 **n** + **n**, and 75 **ŋ** + **n**. Meanwhile, 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 35,907 nouns. These figures are displayed in table 11.

The sets in table 12 constitute an exhaustive list of homophones found in the Sejong Project corpus due to word-internal nasal assimilation. Discounting four probable spelling errors, there are ten homophonic sets out of 2,293 neutralized nouns in the list of 35,907 nouns. There are 732 homophonic tokens out of 1,234,323.

Hwang (2008) investigates the amount of potential noun homophony due to the word boundary process. Her results in table 13 show that a full 51% of all nouns (17,763 out of 35,907) are potentially subject to neutralizing nasal assimilation (**P** = any labial obstruent, **T** = any coronal obstruent, **K** = any dorsal obstruent). 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. Her results are displayed in table 14.

Hwang is careful to point out that these numerical results do not show the actual rate of homophonic usage. Rather, these totals merely indicate the *potential* for derived homophony. Crucially, any noun that might be subject to derived homophony as a consequence of word-final 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

| Set number | Homophonous value | Non-neutralized allomorphs | Word number | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 1 | tʃʌŋni | tʃʌŋ-ɾi | 1 | 정리 | 287 | arrangement |
| | | tʃʌk-ɾi | 2 | 적리 | 3 | dysentery |
| 2 | tʃʌŋnip˥ | tʃʌŋ-ɾip | 3 | 정립 | 71 | triangular position |
| | | tʃʌk-ɾip | 4 | 적립 | 4 | accumulation |
| 3 | kɑŋnon | kɑŋ-ɾon | 5 | 강론 | 8 | sermon |
| | | kɑk-ɾon | 6 | 각론 | 4 | detailed exposition |
| 4 | kudʒiŋnan | ku-tʃik-nan | 7 | 구직난 | 2 | unemployment problem |
| | | ku-tʃik-ɾan | 8 | 구직란 | 2 | unknown |
| 5 | kʲʌŋni | kʲʌŋ-ɾi | 9 | 경리 | 14 | unknown |
| | | kʲʌk-ɾi | 10 | 격리 | 37 | quarantine |
| 6 | pɛŋnʲʌn | pɛk-ɾʲʌn | 11 | 백련 | 15 | white lotus |
| | | pɛk-nʲʌn | 12 | 백년 | 5 | century |
| 7 | pʲʌŋnʲʌk˥ | pʲʌŋ-ɾʲʌk | 13 | 병력 | 90 | replacement depot |
| | | pʲʌk-ɾʲʌk | 14 | 벽력 | 4 | thunder and lightning |
| 8 | jɑŋnʲʌk˥ | jɑŋ-ɾʲʌk | 15 | 양력 | 26 | solar calendar |
| | | jɑk-ɾʲʌk | 16 | 약력 | 3 | vita |
| 6 actual sets (2 due to spelling errors); 12 words; 520 tokens | | | | | | |

*Table 10*
Homophonic noun sets due to liquid nasalization

constitutes about 10% of the lexicon. This suggests that the likelihood of derived homophony in the present case is probably about one-tenth of 2.8%, or .28%. This result is fully in keeping with the remarkably low levels of derived homophony found elsewhere.

| Sequence | Number of non-derived sequences | Number of derived sequences | Homophonic sets |
|---|---|---|---|
| **m**+**m** | 127 | 46 | 0 |
| **n**+**m** | 473 | 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 35,907 nouns 10 homophonic sets | | | |

*Table 11*
Number of nouns subject to nasal assimilation neutralization

1.5 *Coronal assibilation induces very little homophony*

Coronal obstruents assibilate before **s** and **s'**, resulting in (neutralized) **s'** (Martin 1992). Out of the 35,907 nouns in the corpus, 131 words possess the relevant values: 41 lexical, 9 derived. There is a total of one homophonic set as a result of this alternation, 14 tokens in all, as shown in table 15. Out of 1,234,323 noun tokens, 14 are homophonous due to coronal assibilation.

1.6 *Cluster reinforcement induces no homophony*

When a non-aspirated obstruent comes to follow another obstruent, the second value tenses (Kim-Renaud 1975, Martin 1992); this process is called cluster reinforcement. There are 4,048 nouns in the corpus that possess word-medial tensed obstruents: 449 **p'**, 476 **t'**, 1017 **ʧ'**, 1090 **s'**, and 1016 **k'**. In table 16 I provide a list of the homophonic sets due to cluster reinforcement. All three sets are probable spelling errors. There is thus no homophony as a consequence of cluster reinforcement.

1.7 *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. In theory, this assimilation process, unlike the others discussed herein, has the potential to induce a non-negligible amount of homophony. In table 17 I provide the total number of nouns that are potentially subject to this process (17a) alongside the total number of non-variable (genuinely morphologically-sequenced) values with

| Set number | Hom-ophonous value | Non-neutralized allomorphs | Word number | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 1 | tʃaŋmok˥ | tʃak-mok | 1 | 작목 | 7 | poor night vision |
| | | tʃaŋ-mok | 2 | 장목 | 3 | lumber |
| 2 | tʃaŋmul | tʃak-mul | 3 | 작물 | 35 | crops |
| | | tʃaŋ-mul | 4 | 장물 | 14 | loot |
| 3 | tʃaŋmun | tʃak-mun | 5 | 작문 | 24 | composition |
| | | tʃaŋ-mun | 6 | 장문 | 5 | wide-open gate |
| 4 | tʃaŋnʲʌn | tʃak-nʲʌn | 7 | 작년 | 267 | yesteryear |
| | | tʃaŋ-nʲʌn | 8 | 장년 | 8 | prime of life |
| 5 | tʃʰanmul | tʃʰan-mul | 9 | 찬물 | 26 | cold water |
| | | tʃʰas-mul | 10 | 찻물 | 2 | tea |
| 6 | haŋmun | hak-mun | 11 | 학문 | 234 | scholarship |
| | | haŋ-mun | 12 | 항문 | 9 | anus |
| 7 | koŋmul | kok-mul | 13 | 곡물 | 20 | corn |
| | | koŋ-mul | 14 | 공물 | 3 | tribute |
| 8 | kudʒiŋnan* | ku-tʃik-nan | 15 | 구직난 | 2 | unemployment problem |
| | | ku-tʃik-ran | 16 | 구직란 | 2 | unknown |
| 9 | kunmak˥ | kun-mak | 17 | 군막 | 3 | military tent |
| | | kus-mak | 18 | 굿막 | 3 | miners' hut |
| 10 | oɾɛnman | o-ɾɛn-man | 19 | 오랜만 | 116 | unknown |
| | | o-ɾɛs-man | 20 | 오랫만 | 8 | unknown |
| 11 | pɛŋnʲʌn | pɛk-ɾʲʌn | 21 | 백련 | 15 | white lotus |
| | | pɛk-nʲʌn | 22 | 백년 | 5 | century |
| 12 | pʌnːamu | pʌtʃ-na-mu | 23 | 벗나무 | 5 | cherry tree |
| | | pʌs-na-mu | 24 | 벗나무 | 2 | unknown |
| 13 | sʌŋmul | sʌŋ-mul | 25 | 성물 | 17 | unknown |
| | | sʌk-mul | 26 | 석물 | 5 | stone figures |
| 14 | jaŋmul | jak-mul | 27 | 약물 | 42 | medicinal waters |
| | | jaŋ-mul | 28 | 양물 | 7 | penis |
| 10 actual sets (4 due to spelling errors); 20 words; 732 tokens<br>* = non-unique – already listed with liquid nasalization | | | | | | |

*Table 12*
Homophonic noun sets due to word-internal nasal assimilation

| Labials | | | Coronals | | | Dorsals | | |
|---|---|---|---|---|---|---|---|---|
| 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 35,907 nouns; 51% of all nouns | | | | | | | | |

*Table 13*
Potential word-final nasal-assimilated nouns

| Labials | | | Coronals | | | Dorsals | | |
|---|---|---|---|---|---|---|---|---|
| Value | Number of words | % of total | Value | Number of words | % of total | Value | Number of words | % of total |
| **m#** | 87 | 0.2 | **n#** | 44 | 0.1 | **ŋ#** | 354 | 1.0 |
| **P#** | 89 | 0.2 | **T#** | 59 | 0.1 | **K#** | 355 | 1.0 |
| Total | 176 | 0.4 | Total | 103 | 0.25 | Total | 709 | 2.0 |
| 988 words out of 35,907 nouns; 2.8% of all nouns | | | | | | | | |

*Table 14*
Potential word-final nasal-assimilated homophones

| Set number | Homophonous value | Word number | Non-neutralized allomorph | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 1 | **pis͟'ɑl** | 1 | **pit͡ʃʰ-sɑl** | 빛살 | 12 | light ray |
| | | 2 | **pis͟-sɑl** | 빗살 | 2 | comb teeth |
| 1 set; 2 words; 14 tokens | | | | | | |

*Table 15*
Homophonic noun set due to coronal assibilation

which the variable forms might neutralize (17b). Finally, I provide the total number of potentially homophonic sets (17c).

A sizable number of homophonic sets, 62, is possible here. An important point to keep in mind, however, is that this is a VARIABLE pattern of

| Set number | Homo-phonous value | Non-neutralized allomorphs | Word number | Hangul | Token count | Gloss |
|---|---|---|---|---|---|---|
| 1 | ʧ'ɑk'uŋ | ʧ'ɑk-k'uŋ | 1 | 짝꿍 | 3 | buddy |
| | | ʧ'ɑk-kuŋ | 2 | 짝궁 | 2 | unknown |
| 2 | k'ɑkt'ugi | k'ɑk-tu-ki | 3 | 깍두기 | 7 | radish kimchi |
| | | k'ɑk-t'u-ki | 4 | 깍뚜기 | 4 | unknown |
| 3 | kop'ɛgi | kop-pɛ-ki | 5 | 곱배기 | 3 | unknown |
| | | kop-p'ɛ-ki | 6 | 곱빼기 | 3 | double shot |
| No actual sets (all due to spelling errors); 0 words; 0 tokens | | | | | | |

*Table 16*
Homophonic noun sets due to cluster reinforcement

| (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 | | | |
| 5,860 nouns; 91 potentially homophonic sets | | | | |

*Table 17*
Potential homophony due to variable assimilation

neutralization: sometimes neutralization occurs (more often in casual speech), and sometimes it doesn't (more often in formal speech). Jun (1995), for example, investigates the variable labial-to-dorsal obstruent assimilation pattern using the aerodynamic methodology introduced by Silverman & Jun

(1994). He reports gestural reduction of labials (in **p**+**k** sequences) to occur about 35% of the time in casual speech, and about 15% of the time in formal speech. However, Jun finds that this variable pattern does not vary between discrete values **k:** 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. Many cases of near-neutralization are documented in the phonetics literature, increasingly so as analytic techniques become more sophisticated (e.g. Dinnsen & Charles-Luce 1984 and Charles-Luce 1993 for Catalan; Charles-Luce 1985, Port & O'Dell 1985, and Port & Crawford 1989 for German; Slowiaczek & Dinnsen 1985 for Polish; Pye 1986 for Russian; Warner, Jongman, Sereno & Kemps 2004 for Dutch; Bishop 2007 for Andalusian Spanish).

Among variable processes such as Korean coronal assimilation, it is quite possible that potentially homophonic forms in particular are more likely to maintain their distinctive status than are neutralized heterophones, at least in semantically ambiguous contexts. For example, Charles-Luce (1985) finds that 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, presumably, to neutralized values in heterophones. Comparable findings are reported for Dutch (Warner, Jongman, Sereno & Kemps 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) revealed only a few spelling errors documented in the corpus for words with optionally assimilated sequences. This suggests that such sequences are indeed realized as variably- and partially-assimilated, at least a significant portion of the time. Were these sequences completely neutralized, we might expect a higher number of spelling errors: completely neutralized forms are more likely to be subject to spelling errors, since, for such forms, there is less spur-of-the-moment (while writing or typing) evidence to suggest that their phonetic properties are a consequence of neutralization. Indeed, the documented spelling errors in the Sejong Project corpus are not random, but instead always reflect accurate pronunciations of the (presumably intended) word. (I note here that Kim & Jongman (1996), in their investigation of Korean neutralizing aplosivization, find no evidence of near-neutralization under any circumstances. That is, in the case of aplosivization at least – and unlike coronal and labial assimilation – neutralization is apparently complete.)

| Alternation | Number of nouns, both lexical and derived (out of 35,907) | Number of homophonic sets | Number of homophonic words | Number of homophonic tokens (out of 1,234,323) |
|---|---|---|---|---|
| (1) Aplosivization | 10,138 | 15 | 32 | 6,117 (46,781–40,664) |
| (2) Nasal lateralization | 1,001 | 10 | 20 | 288 |
| (3) Liquid nasalization | 695 | 6 | 12 | 520 |
| (4) Word-internal nasal assimilation | 7,592 | 10 | 20 | 732 |
| (5) Coronal assibilation | 131 | 1 | 2 | 14 |
| (6) Cluster reinforcement | 4,048 | 0 | 0 | 0 |
| (7) Variable assimilation | (5,860) | (91) | | (undetermined) |
| Totals | 23,605 | 42 | 86 | 7,671 |

*Table 18*
Summary of neutralizing alternations and derived homophony

### 1.8 *Summary of neutralizing alternations in Korean*

We have now investigated six categorical neutralizing alternations and one variable/gradient neutralizing alternation in Korean. I summarize the numerical results in table 18. There are 42 homophonic sets out of 35,907 nouns involving 86 words. 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.

## 2. 'A VERY LOW LEVEL OF HOMOPHONY' COMPARED TO WHAT?

As pointed out by Surendran & Niyogi (2006), totals such as those presented here acquire greater significance when they are compared to values that do not merge (or, for present purposes, neutralize). In this section I specifically do not compare actual patterns of derived homophony to those that might be encountered assuming raw consonant and vowel combinatorics. Since

languages never possess unconstrained combinations of consonants and vowels, such a comparison would not provide a realistic basis for comparison. Rather, in this section I compare actual patterns of neutralization (along with their amounts of derived homophony) to hypothetical patterns of neutralization (along with their amounts of derived homophony). Two of the three patterns I consider, although plausible and within the realm of possibility, are in fact unattested in Korean. I will examine, in turn, true word-final aplosivization versus hypothetical word-initial laryngeal neutralization, true nasal lateralization versus lateral nasalization hypothetical, and true stop + nasal assimilation versus hypothetical nasal + stop assimilation. The first of these hypothetical cases is an implausible development, but the second and third are quite plausible indeed. It turns out that these hypothetical patterns of neutralization would derive a very large number of homophonic sets, in comparison to the meager amount of actual derived homophony. These results are thus consistent with the hypothesis that patterns of neutralization are more likely to be tolerated when they do not result in a significant increase in the amount of derived homophony. Again, I reiterate that the goal of the present investigation is not to zero in on a genuine upper limit for derived homophony. Rather, the present investigation is merely a single case study.

Our first comparison is between true neutralizing aplosivization versus hypothetical word-initial laryngeal neutralization. In table 19 I compare the number of words and the number of minimally distinctive sets, in both word-initial and word-final position. It is quite clear from these figures that the functional load of word-initial obstruent oppositions is far heavier than that of their word-final counterparts: 26,125 nouns begin with obstruents out of 35,907; 1,912 sets would be homophones due to laryngeal neutralization. By contrast, in reality, 5,134 words end with obstruents, and 15 homophonic sets arise due to aplosivization. This should not be surprising. Recall my proposal that languages do not evolve towards a state in which derived homophony is excessive. Initial laryngeal neutralization among Korean root-initial plosives would induce just such an excessive amount of homophony.

On the one hand, it might be argued that comparing genuine aplosivization to hypothetical word-initial neutralization heightens the significance of the present findings. Languages naturally evolve toward states that exploit the acoustic/auditory space to good advantage, and so lexicons are naturally structured such that more words are distinguished in contexts that are capable of saliently encoding critically distinctive cues, for example, lexically prevocalic contexts over lexically non-prevocalic ones.

On the other hand, it might be objected that, since languages never neutralize consonantal laryngeal distinctions in **#CV** contexts, any comparison to this hypothetical pattern is unilluminating. Consequently, in the following subsections I consider two more plausible hypothetical cases of neutralization. In each case, the actual amount of derived homophony is far lower than the would-be amount.

| Labials | | Coronals | | Dorsals | |
|---|---|---|---|---|---|
| True aplosivization | Would-be word-initial | True aplosivization | Would-be word-initial | True aplosivization | Would-be word-initial |
| 1,202 words | 4,323 words | 408 words | 15,988 words | 3,524 words | 5,814 words |
| 4 sets | 100 sets | 10 sets | 1,755 sets | 1 set | 57 sets |
| 15 true sets of homophones due to aplosivization versus 1,912 sets of would-be word-initial homophones due to laryngeal neutralization | | | | | |

*Table 19*
Genuine sets of homophones due to word-final aplosivization versus would-be homophonic noun sets due to word-initial laryngeal neutralization

| True homophonic sets | Would-be sequence | Number of would-be sequences | Would-be homophonic sets |
|---|---|---|---|
| 10 | **nː** | 681 (from **n** + **ɾ**) | 47 |
| | | 316 (from **l** + **ɾ**) | |
| | | 4 (from **l** + **n**) | |
| | | 187 (lexical **n** + **n**) | |
| 10 true sets due to nasal lateralization versus 47 sets of would-be homophones due to lateral nasalization | | | |

*Table 20*
Genuine sets of homophones due to nasal lateralization versus would-be homophonic noun sets due to lateral nasalization

In table 20 I compare cases of true nasal lateralization to hypothetical lateral nasalization. In comparison to the ten actual homophonic sets due to nasal lateralization, there are 47 would-be homophonic sets due to lateral nasalization.

A comparable result is obtained when comparing true stop + nasal assimilated sequences to hypothetical nasal + stop assimilated sequences: there are 10 sets of homophones due to stop + nasal assimilation, while there would be 1100 sets of homophones due to nasal + stop assimilation, 1072 if we discount unattested + **ŋ** forms (table 21).

To summarize the results of this section, the six categorical cases of neutralization found in Korean yield a total of 42 sets of homophones. By comparison, the three hypothetical cases of neutralization that are not found in Korean would produce 3,059 homophonic sets.

## 3. A brief history of Korean neutralization

Korean 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 Kolye Dynasty, about 600–1000 years ago – which served to supplant a significant portion of its native vocabulary, particularly its noun inventory. During the era of borrowing, Chinese possessed a very limited set of consonants in root-final position: **p˺ t˺ k˺ m n ŋ**. The Korean of this era had these endings, though Middle Chinese final **t** was typically – and oddly – incorporated into Korean as **l**. This is odd since, as stated, Native Korean indeed possessed **t**-final words. However, **t**-final Sino-Korean words were usually written with **lʔ**, in an apparent effort to call attention to their phonetically checked nature, at least in prescriptive pronunciations. Martin

| True lexical N + N sequences | Would-be sequence | Would-be N + N sequences derived from nasal + stop | True N + N sequences derived stop + nasal |
|---|---|---|---|
| 127 | **m + m** | 512 | 46 |
| 473 | **n + m** | 1,179 | 102 |
| 459 | **ŋ + m** | 1,236 | 271 |
| 97 | **m + n** | 911 | 18 |
| 184 | **n + n** | 3,333 | 47 |
| 394 | **ŋ + n** | 4,571 | 75 |
| – | **m + ŋ** | 249 | – |
| – | **n + ŋ** | 1,238 | – |
| – | **ŋ + ŋ** | 1,327 | – |

*Table 21*
Would-be nasal + stop word-internal nasal assimilation sequences versus true stop + nasal word-internal nasal assimilation sequences

(1997) plausibly suggests that the sound in question was borrowed as a tap (**ɾ**). The dearth of roots with final **t** even in Middle Korean – only forty-eight verbs and a few nouns, according to Martin (1997) – suggests a frequency effect: the loan pattern may have been influenced by the infrequency of Native Korean **t**-final roots, and the high frequency of Native Korean **ɾ/l**-final roots. Though nothing crucial hinges on this issue, and though the question is far from resolved (indeed, I am doing Martin an injustice by simplifying his remarkably detailed discussion), I operate herein under the assumption that Middle Chinese root-final **t** was borrowed as **ɾ**, and alternated with **l** when non-prevocalic, much as this value does today.

In addition to these simple stop endings, Korean roots possessed what Martin (1992) terms 'overstuffed' endings, among them **pʰ pt ptʰ ps pst psk pts tsʰ tʰ sp st sn sk kʰ ks lk lp lpʰ lm nts nʰ** (the affricates were not yet palatalized; they were dental) (Martin 1992, Sohn 1999, Song 2005). Martin believes that up to 600 years ago, these endings were realized intact when a vowel followed, but by 400 years ago they had been reduced to a single member when non-prevocalic. Contemporary Korean root-final tensed values seem to be the reflexes of at least some of these previously complex root-final clusters; other root-final clusters survive to the present, including **ps ls ks ntʃ ltʰ lh nh lk lp lpʰ** (Ko 2006). Also around 600 years ago, root-final **ts tsʰ s** began to reduce to **s** when non-prevocalic.

By 400 years ago, non-prevocalic **s** (which was now an alternant of **ts ts$^{h}$ s**) had succumbed to aplosivation, neutralizing to **t**$^{\neg}$. However, Manchu and Mongol loanwords were written with a final **s** as recently as 200 years ago. In the case of final **h**, it was dropped word-finally, but realized as post-aspiration when a lexical stop followed. In Contemporary Korean, root-final **h** has been completely eliminated among noun roots, surviving only in some (morphologically complex) forms.

The huge influx of Sino-Korean nouns was largely coincident with the onset of the process of root-final consonant attrition. What seems to have happened is that this property came to influence Korean phonotactics at large. The gradual attrition of the Korean system of root-final values thus appears to have been set in motion by the influx of Sino-Korean forms.

The trend towards simplification of the Korean coda system has continued down to recent times. Consider some additional findings, which demonstrate that root-final values in Korean are actually merging (in addition to neutralizing).

First, Ito (2006) observes that since the Middle Korean period, ten (out of 13) **t**-final roots have changed to **s**. She further reports that polysyllabic nouns were much more likely to change to **s** than were monosyllabic nouns.

Second, according to Jun (2007), young Korean speakers are introducing a significant amount of variation in the phonetic values among prevocalic root-final obstruents, though with preferences for some values over others. Among roots that heretofore have contrasted **t t$^{h}$ s ʧ ʧ$^{h}$** prevocalically, there is now a substantial amount of variation among these values. In other words, the distinction among these prevocalic root-final values appears to be breaking down. Simplifying somewhat, among the coronals, **s** is the favored variant; among the labials (**p p$^{h}$**), **p** is the favored variant; among the dorsals (**k k$^{h}$ k**'), **k** is the favored variant. These favored variants are also the most frequent in terms of the number of words that possess them. Jun suggests that this is no coincidence: the token prevalence of some variants over others is arguably a consequence of the lexical prevalence of some values over others.

Third, Kang (2006) observes that those coronal-final roots that vary toward **s** tend to be low-frequency items.

Fourth, Albright (2008) shows that these root-final mergers in Korean actually cause very little ambiguity, due to statistical asymmetries in the lexicon of Korean, and due to their frequencies of usage.

The upshot is that root-final obstruents in the noun system are becoming increasingly subject not only to neutralization, but to full merger as well, with few counter-functional consequences. Table 22 provides a simplified timeline of this diachrony. Note in particular that root-final consonant clusters are not listed. Many such clusters evolved into tensed obstruents. Consequently, the introduction of root-final tensed values actually involves a

| Labials | Coronals | | | | | | Dorsals | | Commentary |
|---|---|---|---|---|---|---|---|---|---|
| Korean consonantal root endings, lexically prevocalic, >600 years ago (*italic* values are also Sino-Korean): | | | | | | | | | |
| ***p*** | **$p^h$** | **t** | **$t^h$** | **ts** | **$ts^h$** | **s** | **(h)** | ***k*** | **$k^h$** | Chinese compounds enter<br>Aplosivization expands |
| Korean consonantal root endings, lexically non-prevocalic, >600 years ago: | | | | | | | | | |
| **$p^\urcorner$** | | **$t^\urcorner$** | | **s** | | | **∅** | **$k^\urcorner$** | | Nasal assimilation begins |
| Korean consonantal root endings, lexically prevocalic, >400 years ago: | | | | | | | | | |
| **p** | **$p^h$** | **t** | **$t^h$** | **ts** | **$ts^h$** | **s** | **(h)** | **k** | **$k^h$** | Nasal lateralization begins<br>Liquid nasalization begins(?) |
| Korean consonantal root endings, lexically non-prevocalic, >400 years ago: | | | | | | | | | |
| **$p^\urcorner$** | | **$t^\urcorner$** | | | | | | **$k^\urcorner$** | | **s→$t^\urcorner$** |
| Contemporary Korean consonantal root endings, lexically prevocalic (bold-edged cells may vary; *italic* values are preferred variants; tensed values arose from a collapse of 'overstuffed' root endings, thus simplifying the system – Martin 1992): | | | | | | | | | | | | |
| ***p*** | **$p^h$** | **t** | **$t^h$** | **ʧ** | **$ʧ^h$** | ***s*** | **s'** | **(h)** | ***k*** | **$k^h$** | **k'** | Root-final values begin to merge |
| Contemporary Korean consonantal root endings, lexically non-prevocalic: | | | | | | | | | | | | |
| $p^\urcorner$ | | $t^\urcorner$ | | | | | | $k^\urcorner$ | | | | |

*Table 22*
Simplified timeline of Korean obstruent root ending attrition

decrease in the number of root endings, and not the increase that is suggested by the display.

In short, the inventory of root-final values in Korean seems to be in a rather steady – if slow – state of attrition. The current proposal is that this tendency has remained in place due to the absence of functional pressures that might inhibit it, since, as shown in section 1, the attrition does not create a significant amount of homophony.

## 4. Concluding remarks

Despite the very significant extent of neutralization in Korean, the counter-functional consequences are quite meager: out of 35,907 nouns in the Sejong Corpus, there are only 42 sets of homophones as a consequence of the six categorical neutralizing alternations investigated. These findings are consistent with the hypothesis that, even when there is a natural phonetic tendency toward a particular phonetic state (say, non-prevocalic aplosivization among obstruents, or various sorts of neutralizing consonantal assimilations), a given language will most likely not evolve toward that state if excessive homophony were to result, because the very speech tokens that are produced with homophone-inducing neutralized alternants are also the very speech tokens that might confuse listeners. Accordingly, as a natural, passive consequence, such tokens would not be reproduced as these listeners become speakers (Labov 1994, Silverman 2006). Rather, in such a language, there is likely to be a natural, passive maintenance of contrast, for it is exactly those speech tokens that are communicated successfully to listeners that are more likely to take hold and to become conventionalized as the speech norm. Indeed, as Hoenigswald (1960: 80) writes, 'it is entirely conceivable that threatened homonymy should also be considered as counteracting the forces of "articulatory ease" in holding back an entire merger process and in preserving a phonemic contrast even where it is not minimal'. By contrast, in Korean, there may be no functional counter-pressure inhibiting the natural tendency toward neutralization, since the introduction of the neutralizing patterns does not have significant counter-functional consequences, in the form of significantly increasing the amount of derived homophony.

I have not attempted to specify or determine a numerical upper limit for derived homophony. While this value may ultimately be empirically ascertainable by thoroughly investigating many languages, in this paper I have merely shown that, in Korean at least, the amount of derived homophony is remarkably low. The very low levels of homophony in Korean become especially salient by comparing ACTUAL homophonic sets to plausible WOULD-BE homophonic sets. Indeed, the present findings suggest that anti-homophony and Martinet's original formulation of the 'functional load' argument are genuine factors acting on linguistic sound systems.

## REFERENCES

Albright, Adam. 2008. Explaining universal tendencies and language particulars in analogical change. In Jeff Good (ed.), *Linguistic universals and language change*, 144–181. Oxford: Oxford University Press.

Bishop, Jason. 2007. Incomplete neutralization in Eastern Andalusian Spanish: Perceptual consequences of durational differences involved in s-aspiration. *International Congress of Phonetic Sciences* XVI (*ICPhS* 16), 1765–1768.

Bladon, Anthony. 1986. Phonetics for hearers. In Graham McGregor (ed.), *Language for hearers*, 1–24. Oxford: Pergamon Press.

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.

de Boer, Bart. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

Dinnsen, Daniel & Jan Charles-Luce. 1984. Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12, 49–60.

Flemming, Edward. 1995. *Auditory representations in phonology*. Ph.D. dissertation, University of California, Los Angeles. [Published 2002, Outstanding Dissertations in Linguistics series; New York: Routledge.]

Hockett, Charles F. 1955. *A manual of phonology* (Indiana University Publications in Anthropology and Linguistics 11 ). Bloomington, IN: Indiana University.

Hockett, Charles F. 1967. The quantification of functional load. *Word* 23, 320–339.

Hoenigswald, Henry M. 1960. *Language change and linguistic reconstruction*. Chicago: University of Chicago Press.

Hwang, Jennifer. 2008. Neutralizing nasalization in Korean. Ms., San José State University.

Ito, Chiyuki. 2006. Coronal coda distribution in Korean nouns. Ms., Tokyo University of Foreign Studies, Institute for Languages and Cultures of Asia and Africa.

Jun, Jongho. 1995. *A constraint-based analysis of place assimilation typology*. Ph.D. dissertation, University of California, Los Angeles.

Jun, Jongho. 2007. Variation in Korean stem-final obstruents. Ms., Seoul National University.

Kang, Yoonjung. 2006. Neutralizations and variations in Korean verbal paradigms. In Susumu Kuno, Ik-hwan Lee, John Whitman, Joan Maling, Young-Se Kang, Peter Sells & Hyang-Sook Sohn (eds.), *Harvard Studies in Korean Linguistics* 11, 183–196.

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. *Korea consonantal phonology*. Seoul: 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, Berkeley.

Kingston, John. 1990. Articulatory binding. In John Kingston & Mary E. Beckman (eds.), *Between the grammar and physics of speech* (Papers in Laboratory Phonology 1), 406–434. Cambridge: Cambridge University Press.

Ko, Heejeong. 2006. Base–output correspondence in Korean nominal inflection. *Journal of East Asian Linguistics* 15, 195–243.

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/ (24 July 2009).

Lindblom, Björn, Peter MacNeilage & Michael Studdert-Kennedy. 1984. Self-organizing processes and the explanation of language universals. In Brian Butterworth, Bernard Comrie & Östen Dahl (eds.), *Explanations for language universals*, 181–203. Berlin: Walter de Gruyter.

Lombardi, Linda. 1991. *Laryngeal features and laryngeal neutralization*. Ph.D. dissertation, University of Massachusetts, Amherst. [Published 1994, Outstanding Dissertations in Linguistics series; New York: Garland.]

Martin, Samuel E. 1992. *A reference grammar of Korean*. Rutland, VT: Charles E. Tuttle.
Martin, Samuel E. 1997. How did Korean get *-l* for Middle Chinese words ending in *-t*? *Journal of East Asian Linguistics* 6, 262–271.
Martinet, André. 1952. Function, structure, and sound change. *Word* 8.2, 1–32.
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, Susan. 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, University of California, Los Angeles. [Published 1997, 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.
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. Ms., University of California, Los Angeles.
Steriade, Donca. 1997. Phonetics in phonology: The case of laryngeal neutralization. Ms., University of California, Los Angeles.
Steriade, Donca. 2000. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *Perception in phonology*, 219–250. New York: Academic Press.
Surendran, Dinoj & Partha Niyogi. 2003. *Measuring the usefulness* (*functional load*) *of phonological contrasts* (Technical Report TR-2003-12). Chicago: 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 Nedergaard Thomsen (ed.), *Competing models of linguistic change: Evolution and beyond*, 43–58. Amsterdam & Philadelphia: John Benjamins.
Trubetzkoy, 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.
Wright, Richard. 2004. A review of perceptual cues and cue robustness. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 34–57. Cambridge: Cambridge University Press.

*Author's address:* *Department of Linguistics and Language Development,*
*San José State University, One Washington Square, San José,*
*CA 95192-0093, USA*
*daniel.silverman@sjsu.edu*