

# Frequency and Relexicalization of Clustered Codas in Korean

YOUN-JEOUNG CHOI

*The University of Texas at Arlington*

## 0. Introduction

In Korean, when a morpheme contains a consonant cluster in its underlying form (such as /salm/ ‘life’, /talk/ ‘chicken’), we observe two outcomes in surface representations. In cases when a morpheme containing a consonant cluster in its coda is followed by either another consonant or by a word boundary, the cluster is simplified by deleting one member of the cluster: /salm + to/ ‘life also’ → [samdo], /talk/ ‘chicken’ → [tak]. If, however, a vowel-initial morpheme follows the underlying cluster, both members of the cluster are licensed by the language’s syllable structure and thus both consonant surface: /salm + i/ ‘life + nominative marker’ → [sal-mi], /eps + i/ ‘not exist + adverbial marker’ → [ep-si].

Contrary to prescriptive accounts in which the resyllabification is claimed to be completely predictable (the second consonant of the clustered codas is assumed to be the onset of the following vowel), as the examples in (1) indicate, observations of contemporary Korean usage suggest that there is variation in terms of how underlying clusters actually surface.

### (1) Morpheme-final clustered coda + *i* (nominative marker)

<u>UR</u>	<u>Prescriptively correct form</u>	<u>Observed phonetic forms</u>	
/talk-i/	[tal-ki]	[ta-ki] ~ [tal-ki]	‘chicken’
/hulk-i/	[hul-ki]	[hu-ki] ~ [hul-ki]	‘soil’
/neks-i/	[nek-si]	[ne-ki] ~ [nek-si]	‘soul’
/saks-i/	[sak-si]	[sa-ki] ~ [sak-si]	‘wage’
/kaps-i/	[kap-si]	[ka-pi] ~ [kap-si]	‘price’
/salm-i/	[sal-mi]	[sa-mi] ~ [sal-mi]	‘life’
/tols-i/	[to-li]	[to-li]	‘anniversary’

In (1), except for the example /tols/, all of the observed forms show variation. When the observed forms that differ from the prescriptively correct forms are reviewed, there is a deleted consonant: it can be the first consonant of the clustered codas as in /talk + i/ → [ta-ki], /hulk + i/ → [hu-ki], or it can be the

second consonant of the clustered codas as in /neks + i/ → [ne-ki], /saks + i/ → [sa-ki]. (Although the example /tols + i/ shows no variation, it is also the case that the second consonant of the clustered is deleted.)

Throughout the previous research (Kim-Renaud 1974, Whitman 1985, Kim 1987, Oh 1994, Sohn 2002), cluster coda-related issues have been studied. In most of the early studies, however, the variational aspect of the issue was not very much discussed since they focused on the prescriptively correct forms without any observation of reality. Some researchers (Kim-Renaud 1974, Oh 1988, 1994, Sohn 2002) have tried to explain the variation in resyllabification by means of optional rule application. For example, Oh (1988, 1994) suggests a Lexical Resyllabification Rule which is optionally applied to the noun stem. She argues that the rule application triggers both of the consonants in the cluster of the noun stem to survive. Without the rule application, the noun stem behaves like a simple coda word because of Consonant Cluster Simplification (one member of the cluster is deleted by the rule). More recently, Sohn (2002) proposed that some cluster coda nouns undergo relexicalization, resulting in the nouns being simple coda nouns. She claims that this relexicalization going on the input forms is considered as evidence of linguistic change based on the idea of input-output correspondence. This study, basically, starts from the same idea as Sohn in the sense that the nouns undergo relexicalization, and the input-output correspondence motivates the process. In addition to the language-internal motivation, this study investigates language-external aspects of the relexicalization such as frequency. In particular, this study examines the relationship between relexicalization and frequency based on the studies of Bybee (1985, 1998) and Phillips (1998).

According to Bybee (1985, 1998) and Phillips (1998), there is a connection between morphological formation and frequency, and the connection starts from the concept of “emergence” suggested by Hopper (1987). They argue that knowledge of language use is highly affected by the stored lexicon, which is developed by means of high frequency (repetition) throughout the previous utterances. In other words, as for the speakers’ productions and understanding of new utterances, the role of linguistic experiences, such as frequency (repetition), plays a more crucial role than abstract grammar.

Bearing in mind the role of frequency in morphological formation, this study analyzes actual productions of the target words (seven monosyllabic nouns having cluster codas) of 55 native speakers of Korean, and tries to characterize the observed variation in the resyllabification of clustered codas in terms of relexicalization. In addition, corpus analysis of Korean is used as a prime tool for measuring frequency. By doing so, this study shows why the relexicalization is related to the frequency of the target words.

## 1. Methods

This study investigates data drawn from two types of research: (1) corpus analysis of Korean (written and spoken materials), and (2) linguistic experiments (analysis of the actual productions of the target words by the subjects).

### 1.1. Corpus Analysis

For the corpus analysis, around 330,000 words (260,330 words from written materials and 74,393 words from spoken materials) were examined. By means of a Korean concordance program (*Kuljabi 2*), each occurrence of the seven target words (*kaps* ‘price’, *neks* ‘soul’, *talk* ‘chicken’, *tols* ‘anniversary’, *salm* ‘life’, *saks* ‘wage’, and *hulk* ‘soil’) in the materials was counted, and the results of the analysis were used as evidence showing the frequency of each target word. The descriptions of the materials are provided in Tables 1 and 2.

Table 1. Description of the written corpus

Title	Genre	Year	Writer	Corpus Size (words)
One Feet Human Mind	Essay	1995	Female	61,000
Father	Novel	1996	Male	44,100
Thorn Fish	Novel	2000	Male	47,430
Inheritance	Novel	2002	Female	61,800
Positive Thinking no. 125-130	Essay	2002	?/contributors	46,000
<b>Total</b>				<b>260,330</b>

Table 2. Description of the spoken corpus

Title	Genre	Year	Corpus Size (words)
Kim Han-Kil Show	Talk show	1994-1995	12,437
For Men	Talk show	1995	5,233
News (MBC & KBS)	News	1998	20,328
Sohn Sook Show	Talk show	1996	10,191
Saturday Show	Show	1995	8,787
Meet at Four	Talk show	1994	17,417
<b>Total</b>			<b>74,393</b>

### 1.2. Linguistic Experiment

#### 1.2.1. Subjects

To examine actual productions of the target words, a total of 55 native speakers of the Seoul dialect of Korean were recruited as subjects. Since I focus on the variation within one dialect, only the standard dialect (a dialect spoken in the Seoul and Kyunggi area) was considered as a target dialect. Regarding the standards for selecting subjects, I considered only those who have lived in the Seoul and Kyunggi area from childhood until the present, and who were educated

in this area. Information about the social background of the subjects is collected through a prepared questionnaire. The content of the questionnaire includes questions about age, sex, level of education, and so on.

### 1.2.2. Materials and Procedures

The target words of this study are seven monosyllabic nouns having morpheme-final consonant clusters: *kaps* ‘price’, *neks* ‘soul’, *talk* ‘chicken’, *tols* ‘anniversary’, *salm* ‘life’, *saks* ‘wage’, and *hulk* ‘soil’. The tasks of the subjects are to produce each target word in two different situations. The first situation is that the subjects are asked to produce the target word + *un* (topic marker) through an “echo game.” In the echo game, the subjects are asked to produce the target words without any orthographic stimuli. The basic idea behind the echo game is to elicit the subjects’ more spontaneous and natural speech productions. The examples in (2) are the basic structure of the sentences used by the researcher and the subjects in this game. The task of the subjects is to produce the sentence A-1 by changing the underlined part after the researcher (in the actual experiment, some dummy words were included to hide the target words).

(2) Test 1: echo game

A. Sentence produced by the researcher:

*ku kos-eynun tal-to pyel-to epseyo*  
 there-at moon-also star-also not exist  
 ‘There is neither moon nor star.’

- |                                  |                                  |
|----------------------------------|----------------------------------|
| 1) <i>oli</i> ‘duck’             | <i>talk</i> ‘chicken’            |
| 2) <i>mole</i> ‘sand’            | <i>hulk</i> ‘soil’               |
| 3) <i>kwuum</i> ‘death’          | <i>salm</i> ‘life’               |
| 4) <i>paykil</i> ‘hundredth day’ | <i>tols</i> ‘the first birthday’ |
| 5) <i>sichey</i> ‘corps’         | <i>nelk</i> ‘soul’               |
| 6) <i>hyuka</i> ‘vacation’       | <i>saks</i> ‘wage’               |
| 7) <i>sangphyo</i> ‘label’       | <i>kaps</i> ‘price’              |

A-1. Sentence produced by the subjects:

*ku kos-ey pyel-un isseyo*  
 there-at star-topic marker exist  
 ‘There is a star.’

- 1) *talk-un* ‘chicken’
- 2) *hulk* ‘soil’
- 3) *salm* ‘life’
- 4) *tols* ‘the first birthday’
- 5) *nelk* ‘soul’
- 6) *saks* ‘wage’
- 7) *kaps* ‘price’

The other situation is that the subjects are asked to read the prepared word list. The word list includes the same seven target words as in the echo game. The subjects' productions made through this task are considered as more like conscious and formal speech production since orthographic stimuli usually affect the behavior of speakers. The word list is provided in (3).

- (3) Test 2: reading word list
- 1) *talk-i* 'chicken-*i* (nominative marker)'
  - 2) *hulk-i* 'soil-*i*'
  - 3) *salm-i* 'life-*i*'
  - 4) *tols-i* 'the first birthday-*i*'
  - 5) *neks-i* 'soul-*i*'
  - 6) *saks-i* 'wage-*i*'
  - 7) *kaps-i* 'price-*i*'

## 2. Data Analysis and Discussion

Based on the two linguistic experiments—the echo game and the reading word list—actual productions of the target words were elicited. In Tables 3 and 4, each target word is represented by two phonetic forms and for each phonetic form, the raw number of occurrences and its relative frequency are provided.

Table 3. Results of echo game

Target word	Phonetic form	No. of occurrences	Relative frequency
<i>salm + un</i> 'life + <i>un</i> '	[sal-mun]	36	65%
	[sa-mun]	19	35%
<i>kaps + un</i> 'price + <i>un</i> '	[kap-sun]	33	60%
	[ka-pun]	22	40%
<i>neks + un</i> 'soul + <i>un</i> '	[nek-sun]	26	47%
	[ne-kun]	29	53%
<i>saks + un</i> 'wage + <i>un</i> '	[sak-sun]	7	13%
	[sa-kun]	48	87%
<i>talk + un</i> 'chicken + <i>un</i> '	[tal-kun]	1	2%
	[ta-kun]	54	98%
<i>hulk + un</i> 'soil + <i>un</i> '	[hul-kun]	0	0%
	[hu-kun]	55	100%
<i>tols + un</i> 'first birthday + <i>un</i> '	[tol-sun]	0	0%
	[to-lun]	55	100%

Table 4. Results of reading word list

Target word	Phonetic form	No. of occurrences	Relative frequency
<i>salm</i> + <i>un</i> ‘life + <i>un</i> ’	[sal-mun]	53	96%
	[sa-mun]	2	4%
<i>neks</i> + <i>un</i> ‘soul + <i>un</i> ’	[nek-sun]	48	87%
	[ne-kun]	7	13%
<i>kaps</i> + <i>un</i> ‘price + <i>un</i> ’	[kap-sun]	41	75%
	[ka-pun]	14	25%
<i>talk</i> + <i>un</i> ‘chicken + <i>un</i> ’	[tal-kun]	25	45%
	[ta-kun]	30	55%
<i>hulk</i> + <i>un</i> ‘soil + <i>un</i> ’	[hul-kun]	19	35%
	[hu-kun]	36	65%
<i>saks</i> + <i>i</i> ‘wage + <i>i</i> ’	[sak-si]	16	22%
	[sa-ki]	39	78%
<i>tols</i> + <i>un</i> ‘first birthday + <i>un</i> ’	[tol-sun]	0	0%
	[to-lun]	55	100%

The results presented in Tables 3 and 4, most of all, indicate that there is variation in the resyllabification. In both of the tests, along with the prescriptively correct forms which syllabify both of the consonants, the variational forms which syllabify only one of the consonants in the cluster are also detected. The only case that shows no variation is the example *tols* ‘first birthday’, which is already relexicalized as a single coda word (now, both of the orthographic forms *tols* and *tol* are used in Korean). Moreover, in most of cases, the variational forms turned out to be the preferred phonetic forms. As a matter of fact, the variational form is not a “variation” anymore, but “linguistic change.” The cluster coda words are changing into single coda words. Another noteworthy point in here is that the echo game shows higher frequency of the variational forms than the reading word list test. Without any exception, each target word’s relative frequency of the variational form gets lowered in the reading word list test (e.g., *hulk*: 100% → 65%, *talk*: 98% → 55%, *saks*: 87% → 78%, *neks*: 53% → 13%, and so on). In other words, the subjects tend to delete one member of the cluster more frequently without any orthographic stimuli. On one hand, this result reveals the fact that register plays a role in the behavior of the subjects. When the subjects are asked to read the word list, they try to produce the test items in a more prescriptively correct way. On the other hand, the results confirm the fact that certain words, such as *saks*, are much less sensitive to the register than the rest of the test items. The majority of the subjects produce /saks + V/ as [sa-kV], and this outcome is relatively consistent compared to the other variational phonetic forms.

Depending on the patterns of the variation, the target words can be summarized in three groups. The first group includes words such as *hulk*, *talk*, *saks*, and *tols*, which behave more like single coda words: in both of the tests,

each target word's variational form is preferred to the prescriptively correct form. Differently from the first group, in words such as *kaps* and *salm*, both members of the cluster tend to be syllabified, although the variational forms are still discovered. In the cases of *kaps* and *salm*, the prescriptively correct forms are consistently dominant in both of the tests. Finally, the word *neks* is differentiated from the other groups. In the case of *neks*, depending on the tests, the result comes out differently. While in the echo game the variational form is preferred, in the reading test the prescriptively correct form is highly preferred. This fact suggests that the word *neks* is sensitive to stylistic variation. Bearing in mind the results from the experiments, the frequency of each target word was examined through the corpus analysis. The results of the analysis are presented in Tables 5 and 6.

Table 5. The result of written corpus analysis

Target word	<i>salm</i>	<i>hulk</i>	<i>kaps</i>	<i>talk</i>	<i>neks</i>	<i>saks</i>	<i>tols</i>
Number of occurrences	133	66	39	24	13	1	0

(Total number of words in the corpus: 260,000)

Table 6. The result of the spoken corpus analysis

Target word	<i>kaps</i>	<i>salm</i>	<i>talk</i>	<i>neks</i>	<i>tols</i>	<i>hulk</i>	<i>saks</i>
Number of occurrences	25	14	3	1	1	0	0

(Total number of words in the corpus: 74,393)

Considering the results, in general, the target items are not high-frequency words. In particular, the words *tols* and *saks* show extremely low frequency. According to the results, except for the word *hulk*, in both of the analyses, the order of frequency among the words appears in a similar way. Basically, the words *salm* and *kaps* show relatively higher frequency than the rest of the target item such as *tols*, *saks*, and *neks*. In the cases of *talk* and *hulk*, while the written corpus presents the words as high-frequency words, in the spoken corpus analysis those two words show low frequency. Although there might be a possibility that characteristics of the corpora affect the results, at this point, further discussion of the corpus is left for future studies. In this analysis, the frequency of the target words refers to the overall frequency summing up both of the corpora. Therefore,

the order of frequency of the target words is summarized as follows: *salm* (200) > *hulk* (66) > *kaps* (64) > *talk* (27) > *neks* (14) > *saks* (1), *tols* (1).

In her study, Bybee (1985) has argued that there is a connection between word frequency and lexical analysis. According to her, since “lexical storage” is based on the memory of speakers accomplished by their language use, word frequency highly affects the process of morphological formation. Also, regarding lexical diffusion of sound change, Phillips (1998:123) suggests that “sound changes which require analysis—whether syntactic, morphological, or phonological—affect the least frequent word first.” As Tables 5 and 6 show, in this analysis the words *tols* and *saks* are discovered as the least frequent words. As a matter of fact, in these two words, the deletion of the consonant *s* is very apparent. Due to the low frequency of the words, the morphological information of the words is getting lost, and because of the loss, the sound changes of the two words are reinforced. Especially in the echo game, the results for the four nouns *kaps*, *neks*, *saks*, and *tols* support the connection between the frequency of the words and the degree of sound change by showing a gradual pattern: while the most frequent word, *kaps*, shows the least “sound change,” the least frequent word, *saks*, shows the most “sound change.” One of the important points with relation to /-s/ cluster codas in Korean is that the consonant *s* is never phonetically realized in coda position. In coda position, it is either deleted or neutralized as *t*. The cases of the four nouns tested here are investigated in this respect. Since the nouns *kaps*, *neks*, *saks*, and *tols* are pronounced as single coda nouns (i.e., [kap], [nek], [sak], and [tol]) in their lexicon, the phonetic realization of *s* through resyllabification by actual language use is considered as a way of reinforcing the status of the consonant *s*. As a result, in the relatively high-frequency nouns such as *salm* and *kaps*, lexical information is applied more clearly with relation to resyllabification. Therefore, less sound change is detected. However, as in the case of *saks* and *tols*, the low frequency affects the lexical analysis such as resyllabification in Korean.

The connection between frequency and relexicalization holds up the cases of /-s/ cluster codas. However, in the cases of *hulk* and *talk*, in spite of their relatively high frequency relative to *salm* and *kaps*, their resyllabification process is much more like the cases of *saks* and *tols*. In particular, the results of the echo game suggest that the two nouns *hulk* and *talk* are already relexicalized as /huk/ and /tak/. Almost no speaker produced the prescriptively correct forms. Even in the reading word list test, the frequency of the prescriptively correct forms is less than 50%. Considering the frequency of the words, this result seems to be exceptional. Not only the low frequency, but also high frequency still causes relexicalization. Regarding the cases of *hulk* and *talk*, the motivation of the relexicalization can be discussed from a somewhat different perspective. When the cluster sequences of the two words are reviewed, differently from the /-s/ sequences, both of the consonants in the cluster, /lk/, are appropriate codas in Korean. In other words, both of them can be phonetically realized in coda positions by deleting one member of the cluster. Although it is extremely rare to pronounce both of the consonants together, in very careful reading, it is also

possible. The point is that there is a difference in the process of the lexical analysis between the clusters /-s/ and /lk/. In the case of /lk/ cluster, the relexicalization is motivated by a physiological reason. In the lexicon, both of the consonants in the cluster /lk/ exist as phonologically and phonetically active segments, while the consonant *s* in the cluster /-s/ does not. Therefore, the sound change from [lk] to [k] occurred in the lexicon based on physiological motivation, and this sound change extended to resyllabification. Even if the /lk/ sequence is resyllabified as a coda and an onset respectively, the physiological movement from liquid to velar stop is still not an optimal sequence of articulation. As for the physiologically motivated sound changes, Phillips (1998) claims that the most frequent words are affected most. Based on the theory of ease of articulation, when a word shows high frequency, it seems reasonable to speculate that the sound change is reinforced forward in an easy way. Unfortunately, since there is no other noun with the /lk/ cluster in its coda, further analysis through frequency comparison is not possible.

Among the seven target words, *salm* is another word in which both of the consonants in its cluster coda are phonetically active: /l/ and /m/ are appropriate codas in Korean. As in the cases of *hulk* and *talk*, the segment /l/ is deleted from the base /salm/ when it is followed by a word boundary or a consonant. Also, in this case, the deletion is considered to be motivated by a physiological reason. Considering the frequency of *salm*, it is expected to be relexicalized as /sam/, like the cases *hulk* and *talk*, but it is not. As the results of experiments show, the subjects tend to resyllabify both of the consonants in the cluster /lm/. One interesting fact in the word *salm* is its morphological background. The noun *salm* originated from the verb *sal-ta* by attaching the nominal suffix *-m*. Therefore, as Kim-Renaud (1979) pointed out, both of the segments have a certain kind of morphological information, and because of that both of the segments show stable status with relation to resyllabification. In fact, the lesser degree of sound change in *salm* can be explained in terms of its different morphological background.

### 3. Conclusion

In Korean, the variation in resyllabification of the cluster coda nouns is considered as an instance of relexicalization: from cluster codas to single codas. With relation to relexicalization, this study suggests that there is a connection between word frequency and the degree of relexicalization. In particular, the nouns with /-s/ clusters show a pattern depending on their frequency. In the process of resyllabification, most frequent word, *kaps*, tends to be syllabified by both of the coda consonants, while the least frequent word, *saks*, tends to be syllabified by only one consonant, *k*. In addition to word frequency, this study proposes that the relexicalization is also dependent on the cluster types. In the cases of *hulk* and *talk*, differently from the cases of /-s/ coda cluster nouns, relexicalization is due to their high frequency. Also, in the noun *salm*, its morphological background plays a role regarding relexicalization.

## References

- Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Philadelphia: John Benjamins.
- Bybee, Joan L. 1998. The emergent lexicon. *CLS* 34.2: The Panels. 421-435.
- Hopper, Paul J. 1987. Emergent grammar. *BLS* 13: 139-157.
- Kim-Renaud, Young-key. 1974. Korean consonantal phonology. Ph.D. diss., University of Hawaii.
- Kim, Kee-ho. 1987. The phonological representation of distinctive features: Korean consonantal phonology. Ph.D. diss., University of Iowa.
- Oh, Mi-ra. 1988. The phonological word in Korean. *Lee Hae-Sook Kyoswu Cengnyun Kinyem Nonmuncip*. 223-246.
- Oh, Mi-ra. 1994. A reanalysis of consonant cluster simplification and s-neutralization. In Y.-k. Kim-Renaud, ed., *Theoretical Issues in Korean Linguistics*, 157-174. Stanford, CA: CSLI.
- Phillips, Betty S. 1998. Word frequency and lexical diffusion in English stress shift. In R. Hogg and L. van Bergon, eds., *Historical Linguistics 1995, volume 2: Germanic Linguistics*, 223-232. Amsterdam: John Benjamins.
- Sohn, Hyang-Sook. 2002. Non-uniformity in dialectal variation and linguistic change. In S.-U. Lee, G. K. Iverson, S.-C. Ahn, and Y.-M. Yoo, eds., *Pathways into Korean Language and Culture*, 411-440. Seoul: Pagijong Press.
- Whitman, John. 1985. Korean clusters. In S. Kuno, J. Whitman, I.-h. Lee, and Y.-s. Kang, eds., *Harvard Studies in Korean Linguistics*, 280-290. Cambridge, MA: Harvard University.

The University of Texas at Arlington  
Department of Linguistics & TESOL  
BOX 19559-Hammond Hall 403  
Arlington, TX 76019-0559

younjeoung@hotmail.com