문학석사학위논문

Phonological trends in Seoul Korean compound tensification

서울 방언 합성어 경음화의 음운론적 경향

2016 년 8 월

서울대학교 대학원

언어학과 언어학전공

김 서 영

Abstract

Phonological trends in Seoul Korean compound tensification

Kim, Seoyoung
Department of Linguistics
The Graduate School
Seoul National University

In a Korean compound composed of two nouns, W_A and W_B, if the initial onset of W_B is a lax obstruent, it often undergoes tensification, as in /tol + tam/ [tol.t*am] 'stone wall'. This tensification process does not occur in every compound. In /tol + kye.tan/ [tol.kye.tan] 'stone steps', the initial onset of W_B /k/ remains lax.

Treatment of "exceptions" differs in generative phonology and the recent phonological research. In generative phonology, rules apply categorically. Many of previous studies attempted to define the application condition of the tensification rule, but exceptional cases were found no matter how rules were defined. Therefore, this tensification phenomenon was considered "unpredictable" under the rule-based framework. However, the recent phonology accepts that the distribution of the exceptions themselves is phonologically patterned (Zuraw 2000). In other words, "exceptional" cases might not be a mere exception and might have phonological

reasons to be a variant. Therefore, not a single factor, but rather the interaction of various factors, both phonological and non-phonological ones, decides the occurrence of tensification as a whole. Among these factors, the present study focuses on phonological ones. Specifically, I will examine what phonological factors and how significantly these factors contribute to the overall applicability of tensification.

Zuraw (2011), the first systematic study on the distribution of Seoul Korean compound tensification, argued that both phonological and non-phonological factors attribute to the overall probability of the tensification. As inspired by Ito's (2014) survey study on Yanbian Korean, I performed a similar survey on twenty-one Seoul Korean speakers employing 304 compound words. In the survey results, I identified if the trends caused by each factor are significant. The trends found to be significant are shown in (1).

(1) Trends in Seoul Korean compound tensification

a. Tensification is more likely with high frequency items.

b. Tensification is more likely when W_A ends with an obstruent, followed by a nasal, a liquid and a vowel, in descending order.

c. Tensification is more likely when W_A ends with a liquid and W_B also begins with a coronal consonant.

d. Tensification is less likely when W_B contains a laryngeally marked consonant.

I also attempted to develop a formal analysis of this phenomenon, in a frame of Optimality Theory (OT). Characterizing the variable pattern of the phenomenon, MaxEnt OT (Hayes & Wilson 2008) was employed. Constraints responsible for the occurrence of tensification are REALIZEMORPHEME ("Morphemes are phonologically realized") and IDENT(tense) ("Correspondent segments in the input and output are identical for the feature [tense]").

Various trends found in the existing words were also formalized into the separate OT constraints. For (1a), following Ito (2014), a constraint *TENSE/LOWFREQUENCY (*T/LF, "No tensification for low frequency items") was adopted. For (1b), regarding the highest tensification rate with W_A obstruent final compounds, it was speculated that post-obstruent tensification, obligatory within a single accentual phrase in Korean (Jun 1993), still plays a significant role at the juncture of a compound. Therefore, a constraint *obs-lax ("No lax obstruent after an obstruent in an accentual phrase") was established. Meanwhile, Ito (2014) mentioned that lenition might militate against the application of tensification. And the different tensification rates after a vowel and a sonorant might result from the different preferences for lenition. In this regard, Kirchner (1998) suggested that the impetus for lenition varies depending on the flanking segment: lenition is the most likely when a vowel precedes a target consonant, less likely when a liquid does, and the least likely when a nasal does. This hierarchy accords with the result of my survey where the tensification rate of the W_A vowel final compounds was significantly lower than that of the W_A liquid final ones, which in turn was lower than that of the W_A nasal final ones. Capturing the different impetus for lenition, or blockage of tensification, among the three sonorous W_A coda types, a constraint for each context was established: *TENSE/VOWEL_ ("After a vowel, no tensification"), *TENSE/LIQUID_ ("After a liquid, no tensification"), *TENSE/NASAL_ ("After a nasal, no tensification").

In Korean native monomorphemic words, a coronal consonant right after a liquid coda is always tense. Regarding the trend (1c), it was assumed that the coronal consonants' greater impetus for tensification after a liquid might mirror this pattern in a monomorpheme. Martin (2011) proved that a morpheme-internal constraint diffuses its weaker effect across morpheme boundaries and termed it "*leakage*" in the sense that a phonotactic generalization somewhat leaks from the tautomorphemic domain to the heteromorphemic ones. Capturing this *leakage* effect at the compound juncture, a constraint *L(+)C ("No sequence of a liquid and a lax coronal") was adopted.

As pointed out in Ito (2014), the trend in (1d) is attributed to laryngeal co-occurrence restrictions. The presence of a laryngeally marked consonant, tense ([+constricted glottis]) or aspirated ([+spread glottis]), blocks tensification. Unlike in Ito (2014) in which the presence of a laryngeally marked consonant in both W_A and W_B considerably lowered the tensification rate, in my study, the presence in W_B only showed this effect. The laryngeal co-occurrence restriction holds only within a stem, at least in the Seoul Korean compound tensification phenomenon. Thus the constraint OCP(stem) ("No co-occurrence of laryngeally marked consonants in a stem") was adopted.

Using Maxent Grammar Tool (Hayes 2009a), I took the survey results as the training data and tried to find the specific weights of those constraints, shown in (2).

(2) Constraint weight obtained

OCP(stem) (1.676), *obs-lax (1.533), REALIZEMORPHEME (1.227),

*TENSE/LIQUID_ (1.102), *TENSE/VOWEL_ (0.983), *L(+)C (0.785),

*T/LF (0.429), IDENT(tense) (0.142), *TENSE/NASAL_ (0.0000006)

This weighted set of the constraints reflects the crucial effects observed in the existing lexicon. I also confirmed that the distribution of the data reproduced by this grammar highly accorded with that of the input data (R^2=0.88). Therefore, it can be concluded that the proposed analysis can explain the various trends observed in the phenomenon.

Keywords: Compound tensification, *saisiot*, phonological trends, Probabilistic OT, Maximum Entropy model

Student Number: 2014-20106

Table of Contents

1. Introduction

In a compound composed of two nouns in Korean, if the initial onset of the second element is a lax obstruent, it often undergoes tensification, as in /san + pul/ [san.p*ul] 'wild fire'. Traditionally, this phenomenon has been analyzed as the result of epenthesis of a mono-segmental compound juncture marker, mostly commonly /t/. This epenthetic /t/ first triggers automatic post-obstruent tensification, where a lax obstruent becomes tensified after an obstruent, and then deletes itself through cluster simplification, as illustrated in Table 1 (Kim 1970, Lee 1972, Kim-Renaud 1974, and Ahn 1985).

UR	/san + pul/
Epenthetic /t/	/san + **t** + pul/
Post obstruent tensification	/san + **t** + p*ul/
Cluster simplification	/san + p*ul/
SF	[san.p*ul]

Table 1

A derivation for the compound /san + pul/ [san.p*ul] 'wild fire'.

In fact, this epenthetic /t/ has been orthographically represented as <s>. Because of this, this epenthetic /t/ itself has been called *saisiot* (*sai* = between; *siot*

= the letter ㅅ(/s/)) and this compound tensification phenomenon has been termed *saisiot* phenomenon (Kim 1970: 1, Chung 1980: 28).

This tensification process, which is assumed to consist of the insertion of the epenthetic /t/ and the automatic application of the next two rules shown in Table 1, does not occur in every compound. In other words, some compounds undergo this tensification process while others do not. Then, what decides whether or not a compound undergoes tensification? To answer this question, in a rule-based framework, many studies attempted to define the conditions of the rule application.

The data typically analyzed by traditional generative phonology shows a categorical pattern. Phonological rules are treated as though they apply whenever the defined conditions are met. Even if variant forms were attested, they were marked as an exception or simply ignored. Under this framework, previous studies established the tensification rule and the application conditions, with an aim of offering a complete explanation on the distribution of compound tensification. It has been widely suggested and agreed upon that this *saisiot* used to act as a genitive marker in Middle Korean, which was affixed to the end of the first component of a phrase when the two components are in a genitive relation (Ramstedt 1939, Ahn 1968, and Kim 1994). Considering this earlier function as a genitive marker, when predicting the occurrence of *saisiot* in Contemporary Korean compounds, most previous studies focused on the morphological, semantic, or syntactic properties of the two nouns involved (Shim 1979, Chung 1980, Im 1981, and Oh 1988). In these studies, this tensification rule is expected to apply whenever the two components are in a certain morphological, semantic or syntactic relation. However, no matter how the conditions were defined, exceptional cases were found. That is, none of the factors suggested in the previous studies clearly divide the compounds into those that can and cannot undergo tensification. Because of this, a number of studies agreed that the occurrence of compound tensification is unpredictable within a rule-based framework (Choi 1955: 688, Chung 1980: 30, and Oh 1987: 35).

In recent phonological research, it has been widely accepted that the distribution of the exceptions themselves is phonologically patterned (patterned exceptionality; Zuraw 2000, 2010: 418). That is, the cases that would have been considered "exceptional" under the rule-based framework might not be a mere exception. For example, under the tensification rule which states that tensification apply to sub-compounds where the two components are in a modification relation (Shim 1979), the sub-compound /tol 'stone' + kye.tan 'steps'/ [tol.**k**ye.tan] 'stone steps', as opposed to the other sub-compound /tol 'stone' + tam 'wall'/ [tol.**t***am] 'stone wall', would have been marked as an exception for not undergoing compound tensification. However, this compound's idiosyncratic behavior of not undergoing tensification might be driven by other factors from various modules, such as etymology, frequency, or number of syllables. It has been often noted in the literature that compound tensification does not normally occur if the second element of the compound is a polysyllabic Sino-Korean word (Lee 1972, Kim 1974, Kim-Renaud 1974, and Chung 1980). The second component *kye.tan* of the compound /tol + kye.tan/ is a Sino-Korean word with the two syllables, while the second component *tam* of the other compound /tol + tam/ is not. Therefore, although the morphological condition forces both of these compounds to undergo tensification, the etymology and the stem length of *kye.tan* block the compound /tol + kye.tan/ from undergoing tensification.

The comparison between [tol.kye.tan] and [tol.t*am] clearly shows that a single (morphological) factor does not completely decide whether the tensification rule applies or not. Rather, various factors, which have been proposed in the previous studies to explain the occurrence of compound tensification, seem to interact in such a way that one factor might override another. Among these factors, the present study focuses on phonological factors. Specifically, I will examine what phonological factors and how significantly these factors contribute to the overall probability of compound tensification.

In particular, there are two previous studies that inspired this study. Zuraw (2011) is the first systematic study on compound tensification, suggesting that both phonological and non-phonological factors can influence the occurrence of compound tensification: branching morphological structure, etymology, frequency, final-segment type of the W_A (the first element of a compound), syllable number of W_A and W_B (the second element of a compound), and the presence of tense consonants. Here, since she collected the data from a dictionary (Kuklip kukə yənkuwən 1999; http://stdweb2.korean.go.kr), the data might not correctly reflect actual pronunciations. Compound tensification variably applies even within a single word (Kim-Renaud 1974: 166). Therefore, although the dictionary pronunciations are transcribed as either with or without W_B initial onset consonant tensification, both forms can appear in real speech situations. Based only on the dictionary pronunciation, variations within words were not captured. Ito (2014) carries out a study on the Yanbian dialect, which is similar to Zuraw's (2011) study on Seoul Korean. Ito (2014) focuses on phonological factors, especially on the OCP (Obligatory Contour Principle; McCarthy 1986) effect. Rather than collecting the data using a dictionary, this study performed a survey on 35 native

Korean speakers. Therefore, the data reflected the actual pronunciations by native Korean speakers more accurately. Based on this data, within-word variations were also captured.

In the present study, as inspired by Ito's (2014) survey study on Yanbian Korean, I performed a similar survey on Seoul Korean speakers. Based on the survey results, I will show that a number of factors significantly influence the overall probability of tensification in compounds: frequency, syllable number, W_A final-segment type, the place of W_B initial onset segment, and the presence of laryngeally marked consonants. Actually, some of these factors are already suggested, or partially mentioned in the previous literature for handling exceptional cases. That is, if a compound does not undergo compound tensification despite meeting the conditions of rule application, it was often solved by some of those factors above. However, this study considers that all these factors simultaneously interact and decide the overall applicability of compound tensification. Therefore, this study first investigates each factor in turn, by identifying if the trends caused by each factor are significant in the survey data. And then, in order to look at how significantly each factor influences the probability of compound tensification as a whole, the survey results are fitted in a mixed effect logistic regression model in R (R Development Core Team 2014).

Another goal of this study is to establish a formal analysis on the Seoul Korean compound tensification phenomenon, in the frame of Optimality Theory (OT; Prince & Smolensky 1993). To achieve this goal, I adopt OT constraints that are responsible for the occurrence and the various trends of compound tensification. In order to capture the variable pattern of compound tensification, the numerical weights are assigned to the constraints in the model of Maximum Entropy Grammar (MaxEnt; Goldwater & Johnson 2003, Hayes & Wilson 2008).

The organization of the rest of this paper is as follows. In Section 2, I first introduce how I perform my own survey for collecting the pronunciation data from Seoul Korean speakers. In Section 3, I will report the survey results and show that various factors influence compound tensification in Seoul Korean. In Section 4, I will present an analysis on Seoul Korean compound tensification employing OT constraints, along with descriptions on how previous studies treated relevant factors. In Section 5, I provide a learning simulation using a MaxEnt learner. In Section 6, I state what limitations remain and then in section 7, I conclude this paper.

2. Survey

For investigating a phonological phenomenon that shows a variable pattern, a large-scale data is necessary. Based on a large-scale data, not only various factors affecting the rate of occurrence of the phenomenon but also the significances of these factors can be found. In this study, a survey was performed in order to collect a large-scale data from Seoul Korean native speakers. This section describes the design and the method of the survey.

2.1. Materials

I first started with the list of the common nouns, presented in *Korean usage frequency* (Kang & Kim 2009). From this list, I selected noun-noun compounds, whose W_B initial onset consonant is a lax obstruent that can undergo tensification. Then I selected some more compound words from a Korean dictionary (Kuklip kukə yənkuwən 1999), in order that the survey items are fairly equally distributed across different criteria. When selecting the survey items from these two sources, the compounds are included only if both components of the compound are native Korean words. As often mentioned in the literature, this is the context where tensification occurs most frequently (Lee 1972: 466, Kim-Renaud 1974: 166-167, Chung 1980: 38, and Labrune 1999: 137). Finally, I excluded the items that are unknown to the author of this study, a native speaker of Seoul Korean. The final version of the survey material consists of 304 noun-noun compounds. The distribution of these 304 compounds according to various criteria is shown in Table 2.

Frequency		Syl #		W_A coda		W_B ons		Laryngeal F	
High (34-433)	75	2	20	Obs	45	p	76	Yes	101
Low (0-33)	229	3	181	Liquid	94	k	68	No	203
		4	91	Nasal	88	t	48		
		5	11	Vowel	77	c	64		
		6	1			s	48		
Total	304		304		304		304		304

Table 2

The table shows the distribution of 304 items according to various following criteria: token frequency, syllable number, W_A final-segment type, W_B initial onset consonant, and the presence of laryngeally marked consonants.

2.2. Procedure and participants

The questionnaire form was composed of the 304 survey items, which are listed in random order in a Google Docs document. In the questionnaire, as shown in the sample one in Table 3, the 304 compound words were presented with the two possible pronunciation forms written in standard Korean orthography, one with and the other without tensifying the W_B initial onset consonant. When presenting the 304 compound words, the two components of each compound were given in separate parentheses with the symbol of the morpheme boundary '+' in between, as shown in Table 3. If the two components had been presented as an already compounded word without any explicit boundary in between, *saisiot* should have been marked as a coda <s> in some compounds' W_A final syllable, in compliance with the Seoul Korean orthographic regulation (e.g. *<coŋ.i.co.kak>, <coŋ.is.co.kak>) (Kuklip kukə yənkuwən 1999). If this coda <s> is marked in the questionnaire, participants might have applied tensification more readily than if <s> is not indicated. Thus, in order to rule out the possibility that the participants are affected by the orthographic representation, the two components of each compound were presented separately.

[종이 (coŋ.i)[1]] + [조각 (co.kak)]	
	종이조각 (coŋ.i.co.kak)
	종이쪼각 (coŋ.i.c*o.kak)

Table 3

A sample questionnaire.

Twenty-one native Seoul Korean speakers, who were born and raised in Seoul, participated in this survey. They are all undergraduate or graduate students, whose ages ranged from 21 to 36, and were mostly recruited from the two universities in Seoul (Seoul National University and Yonsei University). The link for the Google Docs document was sent to these participants. They were asked to indicate on the document whether the initial onset segment of the second component is

[1] The phonetic transcriptions in the parentheses '()' were not given in the original questionnaire.

pronounced as tense or lax, when making a compound with the two given components.

3. Trends in existing words

A number of factors cause various trends in the Seoul Korean compound tensification phenomenon. In this chapter, based on the survey results, I will examine a number of factors that are considered or expected to influence the applicability of compound tensification. For doing this, I will identify if the trends caused by each factor are statistically significant in this phenomenon by conducting a linear regression analysis or a Pearson's χ^2 test. I will first report that the probability of tensification differs across compounds and then show various trends found in the survey results in turn.

3.1. Variation

The survey on the twenty-one Seoul Korean native speakers' pronunciations of 304 compound words generated 6,384 datapoints in total. The overall rate of compound tensification in Seoul Korean was 57% (3,644 / 6,384).

In this study, the tensification rate of each compound was calculated, by dividing the number of participants who chose the tensified form by the total number of the participants. For example, fourteen participants chose the tensified form [coŋ.i.c*o.kak] for the compound /coŋ.i + co.kak/ ‘a piece of paper’. The tensification rate of this compound was therefore calculated to be 66.7% (14 / 21).

Figure 1 shows the number of the compounds according to the tensification rate. For example, there were twenty-three compound words that have 100% tensification rate.

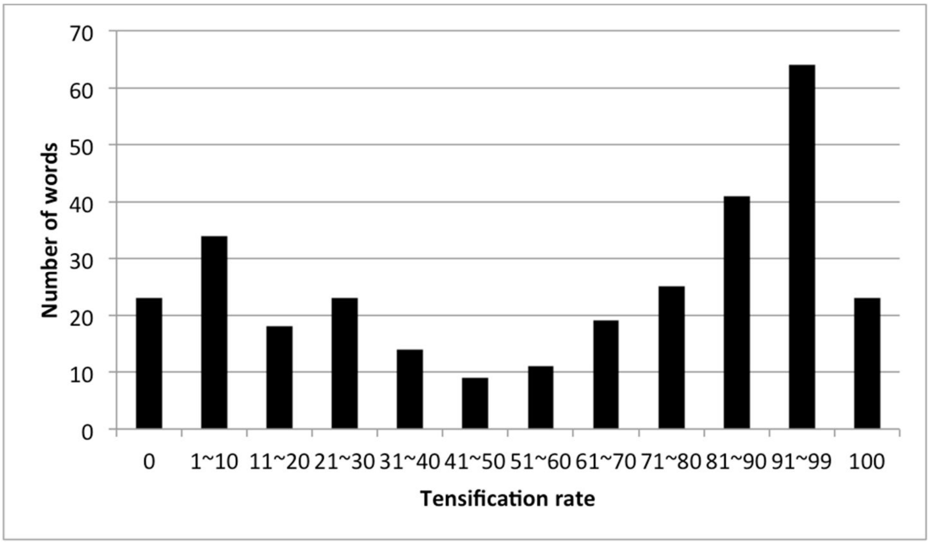

Figure 1

The figure shows the tensification rate on the x-axis (%)
and the total number of compounds on the y-axis.

As shown in the figure above, the preference of tensification differs across the compounds. Some words are easily tensified, whereas others are rarely tensified. Also, some other words are in between, having various tensification rates. This result shows that compound tensification applies with variable probabilities, which are determined by various factors that will be presented in the remainder of this paper.

3.2. Frequency

It was often suggested in the literature that frequently used compound words are more likely to undergo tensification (Oh 1987: 50, Yi 2002: 154, and Lee 2009: 156). In this study, in order to determine how the usage frequency is correlated to the propensity for compound tensification, the usage frequency of each compound word was first taken from *Korean usage frequency* (Kang & Kim 2009). And for each frequency value, the averaged tensification rate was calculated. For example, there were five words whose frequency was twenty-one according to *Korean usage frequency* (Kang & Kim 2009), and the tensification rates of these five words were 95.2%, 95.2%, 90.4%, 80.9%, and 9.5% each. Therefore, the averaged rate for the frequency twenty-one is calculated to be 74.2%.

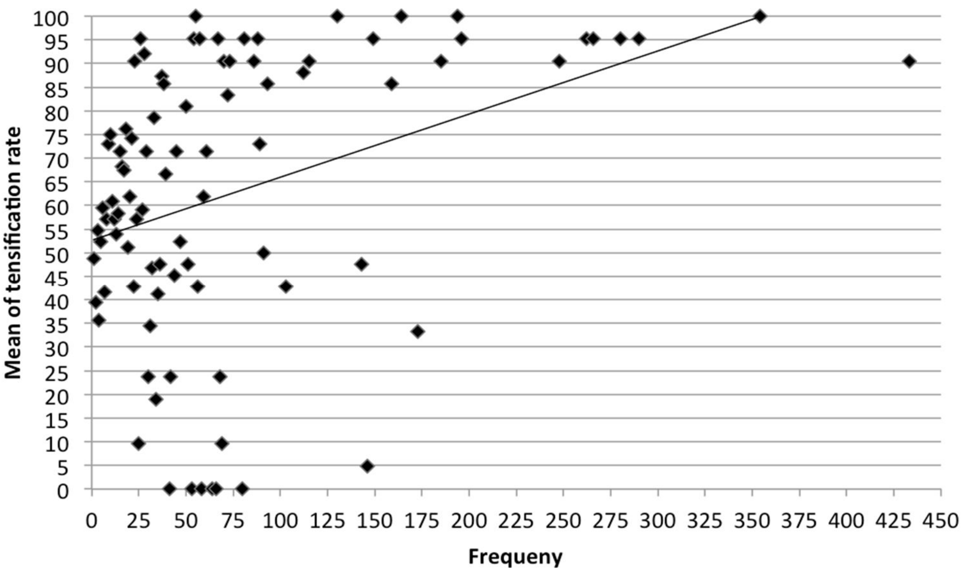

Figure 2

Averaged tensification rate as a function of each usage frequency. The usage frequencies are on the x-axis, and the averaged tensification rates are on the y-axis (%).

Figure 2 shows a positive correlation between the usage frequency and the averaged tensification rate in my data. For example, the tensification rate of the compound /son + pa.tak/ ‘palm’ (frequency: 280) was 95.2% while that of the compound /na.mu + cul.ki/ ‘a trunk of a tree’ (frequency: 14) was only 9.5%. A linear regression analysis with the averaged tensification rate as the dependent variable and the frequency as the independent variable shows a significant result (est. = 0.133, $p < 0.001$). Thus, the usage frequency of a compound significantly affects the probability that tensification applies. The more frequently a compound word is used, the more likely tensification is to apply.

3.3. Syllable number

It was proposed in some earlier literature that polysyllabic stems tend to resist compound tensification. That is, if one or both components of a compound are polysyllabic, tensification is less likely to occur (Lee 1972: 466-467, Kim 1974: 135, Kim-Renaud 1974: 166, and Ahn 1985: 90-91). In order to check if the syllable number of each component indeed affects the probability of compound tensification, the tensification rates according to the syllable number of W_A and W_B are figured.

W_B / W_A	1	≥2
1	73.8 (310 / 420)	60.1 (2010 / 3339)
≥2	54.1 (478 / 882)	48.5 (846 / 1743)

Table 4

Compound tensification rates (%) as a function of the syllable number of W_A and W_B.

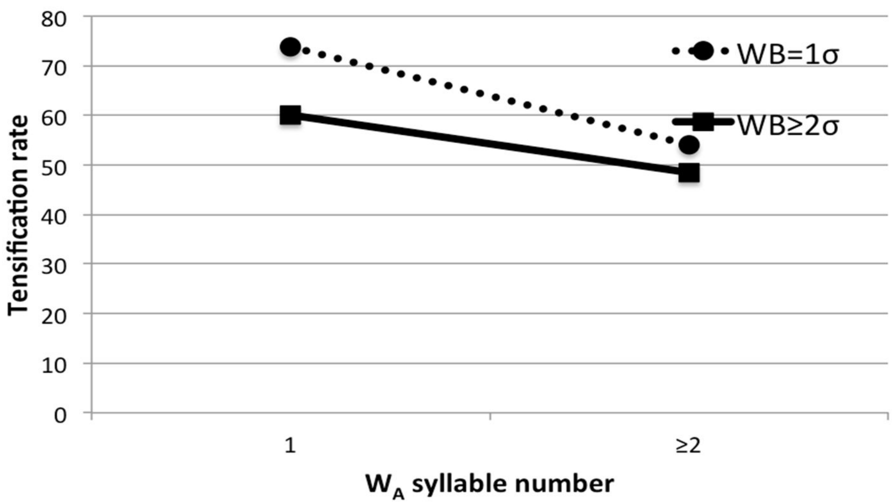

Figure 3

Compound tensification rate as a function of the number of syllables. W_A syllable numbers are on the x-axis, and the tensification rates are on the y-axis (%). The dotted line represents the case where W_B is monosyllabic and the solid line for when W_B is bisyllabic.

As seen in Table 4 and Figure 3, the tensification rate decreases as the syllable number increases from *one* to *more than one* in both W_A and W_B[2]. For example, the tensification rate of the compound /an + kam/ 'lining' was 90.4% while that of the compound /ka.un.te + son.ka.rak/ 'middle finger' was 23.8%. A Pearson's χ^2 test with Yates' continuity correction shows that the differences between *monosyllabic* and *multisyllabic* were significant in all the contexts (when W_A=1σ: $\chi^2 = 28.683$, $p < 0.001$; when W_A≥1σ: $\chi^2 = 7.275$, $p < 0.01$; when W_B=1σ: χ^2 = 44.996, p < 0.00.1; when W_B≥1σ: $\chi^2 = 62.785$, $p < 0.001$). Therefore, the syllable number of W_A and W_B is a significant factor of a compound's propensity to undergo tensification.

3.4. W_A final-segment type

As defined in many of the literature, compound tensification is limited to the cases where W_A ends with a sonorant sound and W_B begins with a lax obstruent (Kim-Renaud 1974: 159-160, Chung 1980: 59, and Ahn 1985: 64). In order to check if the propensity for compound tensification varies among the different sonorant W_A final-segment types, the tensification rate according to each sonorous W_A final-segment type was calculated. Also, for the sake of comparison between sonorant

[2] Among all survey items, there are only four compounds whose W_A is trisyllabic and thus these are aggregated with the compounds whose W_A is bisyllabic. In addition, compounds where W_B is trisyllabic are statistically not different from W_B bisyllabic compounds in tensification ($\chi^2 = 0.875$, $p = 0.349$). In other words, whether the W_B word in a compound consists of two syllables or three syllables does not make a significant difference in a compound's propensity to undergo tensification. Thus, trisyllabic W_B words are aggregated with bisyllabic W_B words. In sum, in both W_A and W_B, the tensification rates were compared only between two groups: *monosyllabic* and *multisyllabic*.

coda and obstruent coda, the tensification rate for W_A obstruent final words was calculated, as shown in Table 5.

Segment type	Tensification rate (%)
Obstruent (45)	86.1 (814 / 945)
Nasal (88)	66.6 (1232 / 1848)
Liquid (94)	47.9 (947 / 1974)
Vowel (77)	40.2 (651 / 1617)

Table 5

Tensification rates (%) as a function of W_A final-segment type.

Table 5 shows the tensification rates according to W_A final-segment type. Despite the fact that post-obstruent tensification, in which a lax obstruent is automatically tensified after an obstruent coda, is obligatory in Korean, the tensification rate after an obstruent coda did not reach 100%. The possible reason for not reaching 100% will be discussed later in §4.3.1. Anyway, the rate after an obstruent coda was still much higher, compared to after the other sonorous codas. Among the other sonorous final-segment types, tensification occurs in the following order, the most frequent to the least frequent: nasal, liquid, and vowel. A Pearson's χ^2 test with Yates' continuity correction shows that obstruent and nasal codas are significantly different in their tensification effects (χ^2 = 120, p = 2.2e-16). Similarly, nasals and liquids are significantly different (χ^2 = 135.3, p = 2.2e-16). It also shows that liquids and vowels are significantly different (χ^2 = 21.1, $p < 0.001$).

Therefore, it can be stated that W_A final-segment type significantly influences the probability of tensification.

3.5. Place of W_B onset consonant

In order to check the interaction between W_A final coda and W_B initial onset, the tensification rates as a function of W_A final segments and W_B initial ones were figured.

W_A Coda / W_B Onset	Obstruent	Nasal	Liquid	Vowel
Coronal	86.7 (510 / 588)	60.2 (557 / 924)	55.4 (571 / 1029)	39.1 (321 / 819)
Non-coronal	85.1 (304 / 357)	73.0 (675 / 924)	39.7 (376 / 945)	41.3 (330 / 798)

Table 6

Tensification rates (%) according to the manner of articulation of W_A final segments and the place of articulation of W_B initial ones.

Table 6 shows the tensification rates according to the manner of articulation of W_A final segments and the place of articulation of W_B initial ones. If W_A ends with a liquid coda, the following coronal onset consonants (*s*, *c* or *t*) are tensified much more than the non-coronal ones (*k* or *p*). For example, the tensification rate of the compound /tol + sot^h/ 'a stone pot' was 66.7% while that of the compound /pal + kup/ 'a hoof' was only 9.5%. Actually, this interaction between W_A final liquid coda and W_B initial coronal onset in native compound words was presumed in Kim (1994: 51) and Ha (2006: 115).

A Pearson's χ^2 test with Yates' continuity correction shows that the probabilities of tensification between coronal and non-coronal consonants after a liquid coda are significantly different from each other ($\chi^2 = 48.037$, $p < 0.001$). Thus, the coronal consonants undergo tensification more frequently than the velar or labial consonants, when a liquid coda precedes it. In other words, the place of W_B initial onset segment itself becomes a significant factor in tensification, at least after a liquid.

3.6. Laryngeal co-occurrence restrictions

In Japanese, there is a well-known morphophonological phenomenon that occurs between two nouns when these two are forming a noun compound, termed Rendaku (Ito & Mester 1986, Vance 1987). Just as the initial onset consonant of the compound's second component undergoes tensification in Korean, in Japanese Rendaku, the initial onset obstruent of the compound's second component undergoes voicing, as in /o.ri/ 'fold' + /ka.mi/ 'paper' [o.ri.**g**a.mi] 'paper folding'.

Regarding Rendaku, it is well known that the voicing is blocked when W_B already contains a voiced obstruent (Lyman 1894: 161). For example, in /hitori/ 'alone' + /tabi/ 'travel' [hi.to.ri.**t**a.bi] 'traveling alone', the voicing of the initial onset of W_B (/t/) is suppressed, because the W_B stem already contains a voiced segment /b/. This phenomenon has been formally analyzed through the OCP effect or co-occurrence restrictions: a co-occurrence of voiced obstruents is not allowed within a stem in Japanese.

In Japanese, the compound marking process, which basically involves the occurrence of another voiced obstruent, can be suppressed by the OCP constraint that bans a co-occurrence of voiced obstruents. Given that Rendaku and compound tensification are both junctural phenomena that occur in a compounding process (Labrune 1999), in Korean, the compounding process, which brings about the occurrence of another tense consonant, can be suppressed by the OCP constraint that bans a co-occurrence of tense consonants.

3.6.1. The presence of laryngeally marked consonants

It was already proven in Zuraw (2011) that having a tense consonant in either W_A or W_B prevents tensification to some extent. In other words, a co-occurrence of tense consonants is restricted in Korean compounds. Later, in Ito's (2014) study on Yanbian Korean, many aspects of this OCP effect in compound tensification were intensively examined. She proved that not only a tense but also an aspirated consonant can suppress tensification. Tense and aspirated are both laryngeally marked, with [+constricted glottis] and [+spread glottis], respectively, whereas lax is laryngeally unmarked (Hayes 2009b: 99). Therefore, that the presence of a tense or aspirated blocks compound tensification arises from the restrictions on the co-occurrence of laryngeally marked segments.

Unlike in Zuraw's (2011) study in which the presence of a tense consonant in either W_A or W_B suppresses the compound tensification to similar extent, in this study, the presence of a laryngeally marked consonant in W_A and W_B displays the different degrees of OCP effect: the presence of a tense or aspirated consonant in W_B significantly lowers the tensification rate but the presence of one in W_A does not lower the tensification rate.

#	Type of consonant in W_B	Tensification rate (%)
0	Lax (259)	59.7 (3302 / 5439)
1	Aspirated (17)	31.1 (104 / 357)
	Tense (17)	31.3 (112 / 357)

Table 7

Tensification rates (%) according to the type of the consonant included in W_B.
The symbol '#' indicates the number of the laryngeally marked consonant included in W_B.

Table 7 shows the tensification rates according to the phonation type of the consonant included in W_B. In the category *Tense*, the compounds containing one tense consonant in W_B are included. Not only the underlyingly tense consonants but also the underlyingly lax consonants after an obstruent coda are counted as tense, considering that post-obstruent tensification is obligatory in Korean simplex words. In the category *Aspirated*, the compounds containing one aspirated consonant in W_B are included. The *Lax* category is composed of the compounds that contain neither a tense nor aspirated consonant in W_B. In other words, this category includes the compounds that only contain lax or sonorant consonants in

W_B. Also, underlyingly aspirated consonants at the end of W_B were considered lax since they would surface as unreleased lax counterparts through the automatic coda neutralization when W_B is produced in isolation. That is, if an aspirated consonant in a word-final coda position is the only aspirated consonant contained in W_B, this compound belongs to the *Lax* category. In fact, in Ito (2014), underlyingly aspirated word-final codas in W_B are counted as aspirated, since it may be realized as an onset in inflected forms and hence released. However, in this study, consonant types are determined based on the independent pronunciation form of the compound word and therefore the aspirated word-final codas are considered lax. Actually, the general patterns remained the same even when they were counted as aspirated anyway (*Lax* 60.1%, *Aspirated* 31.3%, and *Tense* 31.1%).

In addition, in order to examine the OCP effect in W_B more systematically, I excluded from analysis the test words in which W_B has no consonants other than the initial onset, specifically eleven words with monosyllabic codaless W_B (e.g. /coŋ.i + pæ/ 'paper boat' and /an.kæ + pi/ 'misty rain'). This is why the total number of the items included in the three categories in Table 7 is 293, not 304.

As seen in Table 7, the presence of a tense or aspirated consonant in W_B significantly lowers the tensification rate. Compared to the tensification rate of the compounds that contain no laryngeally marked consonant, the rates of those that contain a tense or aspirated consonant decreased by half. A Pearson's χ^2 test with Yates' continuity correction shows that aspirated *vs.* lax, and tense *vs.* lax in W_B are significantly different in their effects on tensification (χ^2 = 117.9, p < 0.001; χ^2 = 135.5, p < 0.001), whereas aspirated *vs.* tense is not (χ^2 = 0.325, p = 0.568). Therefore, the presence of a laryngeally marked consonant in W_B, tense or aspirated, shows a clear OCP effect.

Whereas the tensification rate substantially drops when there is a laryngeally marked consonant in W_B, the presence of one in W_A does not show any OCP effect, as shown in Table 8.

#	Type of consonant in W_A	Tensification rate (%)
0	Lax (230)	58.1 (2809 / 4830)
1	Tense (40)	62 (521 / 840)
	Aspirated (26)	56.4 (308 / 546)
2	Tense and aspirated (4)	4.7 (4 / 84)
	Two tenses (4)	2.3 (2 / 84)

Table 8

Tensification rates (%) as the type of consonants included in W_A.
The symbol '#' indicates the number of the laryngeally marked consonants included in W_A.

Table 8 shows the tensification rates according to the type of the consonant included in W_A. When compounds are categorized to *Tense*, *Aspirated*, and *Lax*, basically the same criteria from Table 7 was applied. However, a new criterion was added for some compounds with the cluster /lk/ at the end of W_A. Considering that cluster simplification and coda neutralization both apply to this cluster, it was counted as lax consonant. No matter which part of the cluster remains, it is either [k] or [l] and thus all /lk/ coda clusters are counted as lax.

As seen in Table 8, the tensification rate of the compounds having a tense consonant in W_A is rather higher than the rate of those having only lax consonants. A Pearson's χ^2 test with Yates' continuity correction shows that a tense consonant and lax consonants are significantly different in their effect on tensification (χ^2 = 4.255, $p < 0.05$). Thus it seems that a tense consonant in W_A encourages tensification, which is opposite of the expectation that the presence of a tense consonant would block tensification.

The tensification rate of the compounds having an aspirated consonant in W_A is lower than that of those having only lax consonants. However, the effects of lax and aspirated consonants on tensification were not significantly different (χ^2 = 0.544, $p = 0.46$). In short, both tense and aspirated consonants in W_A have no clear OCP effect: containing a tense consonant rather seems to promote tensification, and containing an aspirated consonant makes no significant difference on tensification compared to only containing lax consonants. These results are different from those of Zuraw (2011), where the presence of a tense consonant in W_A significantly lowered the tensification rate.

As mentioned above, the presence of a laryngeally marked consonant in W_A does not show any OCP effect. However, although it is not appropriate to make a generalization in the present study due to the small sample size, when there are two laryngeally marked consonants in W_A, a dramatic OCP effect seems to emerge. As seen in Table 8, four words containing both tense and aspirated consonants in W_A and the other four words containing two tense consonants in W_A both have very low tensification rates. This dramatic OCP effect seems to be displayed regardless of the type of laryngeal feature because these two groups are proven to be not very different from each other (χ^2 = 0.172, $p = 0.677$). In other words, if two laryngeally marked consonants, either two tenses or one tense and one aspirated, are contained in a compound, the tensification rate suddenly drops.

In the remainder of this paper, no additional analysis specifically for this maximized OCP effect will be included. The compounds having two laryngeally marked consonants are those that are affected by this maximized OCP effect. However, these compounds are only a few of all the survey items (8 out of 304), and as mentioned earlier, this sample size is too small to make a generalization. Therefore, I only propose a conjecture here and leave its investigation for a future study.

In conclusion, the OCP effect in W_A is displayed only when there is more than one laryngeally marked consonant, and the presence of only one marked consonant does not significantly lower the tensification rate.

3.6.2. Locality of laryngeally marked consonant

In the previous sub-section, it was examined whether the presence of a tense or aspirated consonant in W_A and W_B blocks tensification. The result was that one laryngeally marked consonant in W_B substantially blocks tensification, while more than one in W_A is needed to show the OCP effect.

Regarding the distance between the laryngeally marked consonant and the site of tensification, it is unclear in my data whether a laryngeally marked segment adjacent to the site of tensification has a stronger effect in blocking tensification. As mentioned, W_B is where OCP effect occurs. Therefore, if locality effect exists for real, it would only appear in W_B. However, this locality effect cannot be examined in W_B reliably, due to the insufficiency of the relevant items. For example, among the W_B words containing a laryngeally marked consonant, all the instances except for one (/ma.nɨl + caŋ.a.c*i/ ‘pickled garlic’) are local, and hence the comparison between local and non-local cases was impossible. Eventually, the locality effect could only be checked for the tense consonants in W_A, where no OCP effect was found. Although the results did not show any locality effect in W_A, it is unsure whether this is because W_A has no OCP effect in the first place or locality effect actually has no existence.

Investigating the locality effect in W_A, compounds whose W_A is monosyllabic are excluded since they do not have non-local case. In addition, in order to rule out the potential OCP effect, compound words containing two laryngeally marked consonants in W_A and those containing a laryngeally marked consonant in W_B are also excluded.

Consonant type in W_A	Locality	Tensification rate (%)
Tense (18)	Local (12)	51.9 (131 / 252)
	Non-local (6)	58.7 (74 / 126)
Aspirated (5)	Local (4)	59.5 (50 / 84)
	Non-local (1)	85.7 (18 / 21)
Lax (82)		57.3 (988 / 1722)

Table 9

Tensification rates (%) according to the consonant type
and the locality of the laryngeally marked consonant in W_A.

Table 9 shows the tensification rates according to the consonant type and the locality of the laryngeally marked consonant in W_A. The tensification rates for both local and non-local case of each laryngeally marked consonant are presented. The tensification rate for the category *Lax*, which consists of the W_A multisyllabic compounds containing no laryngeally marked consonant, is also presented.

As seen in the table above, there is only one non-local case for aspirated (/c^hən.tuŋ + so.ri/ 'thunder sound'), and thus the locality effect cannot be checked for aspirated consonants in W_A. For the tense consonants, the local cases have a lower tensification rate than the non-local cases but the difference between these two turned out to be not significant (χ^2 = 1.280, p = 0.257). Even when compared to the tensification rate of the compounds in *Lax* category, the rate of the local case was not significantly different (χ^2 = 2.387, p = 0.122).

In conclusion, whether a laryngeally marked consonant in an adjacent syllable has a stronger effect in blocking tensification could only be checked for the tense consonants in W_A. However, in W_A, the tense consonants in an adjacent syllable do not show a significant difference in tensification, compared to those that are not in an adjacent position. Nor were these significantly different from W_A multisyllabic words containing only lax consonants. In other words, no locality effect was found in my data. However, it is not enough to say that the compound tensification phenomenon in Seoul Korean has no such locality effect. This is because it is not sure whether this result stems from the absence of OCP effect in W_A or from the actual absence of locality effect. Therefore, in the remainder of this paper, I will not look at this effect any further. A more reliable investigation on this locality effect can be a goal of future study.

As one of the significant factors affecting Seoul Korean compound tensification, laryngeal co-occurrence restrictions have been examined in this section. Regarding the presence of a laryngeally marked consonant in W_B, a tense and an aspirated consonant both show an obvious effect in suppressing tensification. On the contrary, containing only one laryngeally marked consonant in W_A shows no such effect; a tense consonant rather seems to encourage tensification and an aspirated consonant makes no significant difference from lax consonants. However, for the eight compounds having two laryngeally marked consonants in W_A, a dramatic OCP effect seemed to emerge because tensification hardly ever occurred in these items. Regarding the locality of the laryngeally marked consonant, it could be examined only within the compounds containing a tense consonant in W_A, but no locality effect was found in the data; the local tense items were significantly different neither from the non-local tense items nor from the multisyllabic lax items.

3.7. Mixed effect logistic regression model

So far in this section, I have mentioned the various factors affecting the applicability of compound tensification: frequency, syllable number, W_A final-segment type, the place of W_B onset consonant, and the presence of laryngeally marked consonants. As mentioned earlier in this paper, none of these factors clearly distinguish the compounds that can and cannot undergo tensification. They do affect whether or not compound tensification occurs, but only up to a certain point. All the factors simultaneously interact in a way that one might override another and contribute to the overall probability of tensification. That is, one factor's strong effect could make up for the weak effect of another.

Investigating how significantly each factor influences the probability of compound tensification as a whole and identifying these factors' relative strengths, I constructed a mixed effect logistic regression model using the *glmer* function from the lme4 package (Bates *et al.* 2014) in R (R Development Core Team 2014). The occurrence of compound tensification was the dependent variable (with non-tensified as the reference) and the various factors introduced throughout the previous sections were the independent variables: usage frequency (numerical: *log(frequency+1)* transformed), W_A final-segment type (categorical: backward difference coding utilized, *vowel* > *liquid* > *nasal* > *obstruent*), W_B initial onset place (categorical: *non-coronal* (ref), *coronal*), W_A / W_B syllable number (categorical: *monosyllable* (ref), *multisyllable*), W_A consonant type (categorical: backward difference coding utilized, *lax* > *tense* > *aspirated*[3]), W_B consonant type (categorical: backward difference coding utilized, *lax* > *tense* > *aspirated*). Random intercepts were set for item and subject. In order to check the interaction, W_A final-segment type and W_B initial onset place were sum-coded. Table 10 shows the results.

[3] Again, in W_A consonant type, eight items having two laryngeally marked segments (four compounds with both tense and aspirated and the other four compounds with two tense consonants) are excluded, due to the small sample size.

	Estimate	Std. Error	Z value	P(>|z|)
(Intercept)	- 1.7286	0.639	- 2.701	0.006 **
Frequency	0.9797	0.246	3.970	< 0.001 ***
W_A coda (liquid-vowel)	- 0.2350	0.547	- 0.429	0.667
W_A coda (nasal-liquid)	1.9995	0.477	4.188	< 0.001 ***
W_A coda (obs-nasal)	2.1880	0.676	3.235	0.001 **
W_B onset place (cor)	- 0.0466	0.280	- 0.167	0.867
W_A consonant (tense-lax)	0.0009	0.390	0.002	0.998
W_A consonant (asp-tense)	0.1696	0.575	0.294	0.768
W_B consonant (tense-lax)	- 2.5790	0.585	- 4.406	< 0.001 ***
W_B consonant (asp-tense)	- 0.6172	0.797	- 0.774	0.438
W_A syllable number (multi)	0.4161	0.335	1.241	0.214
W_B syllable number (multi)	- 0.3458	0.370	- 0.934	0.350
W_A coda (L-V): W_B onset (cor)	1.2354	0.712	1.735	0.082
W_A coda (N-L): W_B onset (cor)	- 1.5842	0.665	- 2.380	0.017 *
W_A coda (P-N): W_B onset (cor)	- 0.6750	0.854	- 0.790	0.429

Table 10

Results of a mixed effect logistic regression model. In the last three rows, V stands for vowels, L for liquids, N for nasals, P for obstruents, and *cor* for coronal consonants.

The result in Table 10 can be interpreted as follows.

First, it is confirmed that the higher the frequency of a compound is, the more likely tensification is to apply.

Regarding W_A final-segment types, an obstruent coda tensifies the following onset segment more readily than a nasal coda does, which in turn tensifies more than a liquid coda does. The effect of a liquid coda on tensification is not significantly different from that of a vowel.

With regard to the OCP effect, neither a tense nor an aspirated consonant in W_A was significantly different from a lax or a tense consonant, respectively. In other words, the differences between aspirated, tense and lax consonants are not very significant in W_A, accurately mirroring the survey results in which only one laryngeally marked consonant in W_A displayed no significant OCP effect at all. On the contrary, the probability of tensification becomes much lower when a tense consonant is contained in W_B than when only lax consonants are contained in W_B. An aspirated consonant is not different from a tense consonant in its tensification effect in W_B, correctly reflecting the survey results in which both a tense and an aspirated consonant in W_B significantly dropped the tensification rates by a similar degree.

The syllable number was not a statistically significant factor in this model. Neither the syllable number of W_A nor that of W_B significantly influenced the propensity of tensification.

The place of W_B initial onset consonant by itself is not a significant factor in this model. However, it becomes significant when W_A final coda changes from liquid to nasal. In other words, there is an interaction between the W_A coda and the W_B onset. The result shows that coronal onset consonants are more likely to undergo tensification after a liquid, compared to after a nasal.

3.8. Summary

Tensification occurs variably across compounds, but there are trends within the variation. In this section, a number of the significant trends in the distribution of Seoul Korean compound tensification were introduced. The trends found can be summarized as follows:

(1) Trends found in Seoul Korean compound tensification

a. Tensification is more likely with high frequency items.

b. Tensification is more likely when W_A ends with an obstruent, followed by a nasal, a liquid and a vowel, in descending order.

c. Tensification is more likely when W_A ends with a liquid and W_B also begins with a coronal consonant.

d. Tensification is less likely when W_B contains a laryngeally marked consonant.

As often mentioned in the literature (Oh 1987: 50, Yi 2002: 154, and Lee 2009: 156), more frequently used compounds are indeed more likely to undergo tensification, according to the survey results.

Many of the previous literature defined that compound tensification occurs in the cases where W_A ends with a sonorous segment and W_B begins with a lax obstruent (Kim-Renaud 1974: 159-160, Chung 1980: 59, and Ahn 1985: 64). My survey results indicate that the propensity for compound tensification differs among the different sonorant types. The tensification rate was the lowest with W_A

vowel final compounds, higher with W_A liquid final compounds, and the highest with W_A nasal final compounds. Above all, the W_A obstruent final compounds had the higher tensification rate than all the W_A sonorant final compounds.

According to the survey results, it was found out that the place of articulation of the initial onset consonant of W_B itself becomes a significant factor of compound tensification, at least when a liquid coda precedes it. That is, when W_A ends with a liquid and also when W_B begins with a coronal consonant, compound tensification is more likely to be applied.

Last but not least, the OCP effect in compound tensification was confirmed. In Japanese Rendaku, the compounding process, which basically involves an occurrence of another voiced segment, is suppressed by the preexistence of a voiced segment in W_B. The Korean compound tensification phenomenon showed a similar pattern with the Japanese Rendaku phenomenon. According to my survey results, a compound word is less likely to undergo tensification if W_B word already contains a laryngeally marked consonant. However, the existence of a laryngeally marked segment in W_A displayed no OCP effect of comparable size. These results are different from those of Zuraw (2011), in which the presence of a tense consonant in both W_A and W_B suppresses tensification.

As mentioned, some of these factors found in my survey results have been partially suggested in the literature already. This study considers all these factors as a whole, fitting a regression model to the data with all the factors included as the variables. By doing so, these factors' significances and their relative strengths were identified.

4. Analysis

The previous chapter introduced various trends in Seoul Korean compound tensification. In this chapter, I will try to capture this variable pattern within the frame of Optimality Theory (OT; Prince & Smolensky 1993).

Standard OT is characterized by a set of strictly ranked constraints. However, under a single fixed constraint ranking, phonological variations cannot be captured, since a single output form will be always chosen as a winner. Formalizing the variable pattern of Seoul Korean compound tensification requires a model that can generate more than one candidate as a winner, and also can capture the frequency distribution of attested outputs. To meet these conditions, I will adopt the model of Maximum Entropy Grammar (MaxEnt; Goldwater & Johnson 2003, Hayes & Wilson 2008). In this model, unlike standard OT, constraints are assigned numerical weights, which are a measure of the constraint's importance. Based on the probability distribution of the candidates, this model assigns weights to each constraint, with the goal of maximizing the probability to reproduce the observed pattern in the candidates' distribution. For example, if the candidate distribution shows that one candidate is less frequent than the other, the constraints violated by this infrequent candidate are assigned relatively higher weights.

In this chapter, I will first propose the constraints that are responsible for the occurrence of compound tensification. Also, various trends, which were introduced throughout the previous section, will be formalized by adopting additional constraints. After the constraint set responsible for the pattern of Seoul Korean compound tensification is established, in §5, I will assign the specific weights to each constraint, by using the computer software, Maxent Grammar Tool (Hayes 2009a).

4.1. The occurrence of compound tensification

First of all, constraints responsible for the occurrence of compound tensification need to be established. As mentioned earlier in this paper, in much of the traditional literature, it has been stated that this compound tensification occurs due to an epenthetic segment, mostly commonly /t/, which is orthographically represented as <s> and called *saisiot*. And then an automatic process of post-obstruent tensification and cluster simplification follow (see Table 1). In fact, other than the most commonly claimed /t/, the previous studies suggested various underlying representations of this epenthetic segment marking the juncture.

This study only focuses on the compound tensification phenomenon but there is another phenomenon that is often observed in the other compounding processes: nasal lengthening. When W_A ends with a vowel and W_B begins with a nasal, the nasal lengthens as in /pæ + no.ræ/ [pæ**n**.no.ræ] <pæ**s**.no.ræ> 'sailor's song' (Kim 1970: 11). Since tensification and nasal lengthening both occur at the juncture of a compound and share the same orthographic representation <s>, much of the previous literature, mostly under rule-based phonology, attempted to propose a unified solution to both phenomena. First, as mentioned above, some studies proposed that /t/ is the underlying form for it (Ryu 1963, Kim 1970, Kim-Renaud

1974, Huh 1986, and Ahn 1985). By assuming the epenthesis of /t/, both tensification and nasal lengthening could be explained through a unified set of rules. More specifically, this epenthetic /t/ triggers either tensification or nasal lengthening depending on the following segment. If W_B begins with a lax obstruent, this epenthetic /t/ first triggers post-obstruent tensification and then deletes through cluster simplification. If W_B begins with a nasal, this epenthetic /t/ assimilates to the following nasal segment through an automatic nasal assimilation in which an obstruent is nasalized when a nasal immediately follows (Kim-Renaud 1974: 219). Other studies proposed that the underlying form for the compound juncture marker is /s/, considering that <s> is the orthographic letter for both nasal lengthening and tensification phenomena (Kim 1992, Moon 1997). This epenthetic /s/ can also explain both tensification and nasal lengthening in a similar way that the epenthetic /t/ explains them, since both /s/ and /t/ are neutralized to [t] in coda (Ahn 1985: 171).

However, Chung (1980: 59-63) does not posit a segment insertion. He claimed that if /t/ epenthesis is assumed as in much of the previous literature, it creates otherwise impossible clusters (e.g. *mt*, *nt*, etc.) before the cluster simplification takes place (see /san*t*.pul/ -> /san*t*.p*ul/). He mentioned that this problem of creating nonexistent clusters arises if a segment, whatever it may be, is inserted. Therefore, instead of adopting a rule that inserts a segment such as /t/ or /s/ solely for the purpose of tensifying its next following segment and deletes it right after this purpose is accomplished, he would rather adopt a rule that simply tensifies the following lax consonant. This analysis was based on the assumption that the two phenomena, tensification and nasal lengthening, are to be treated independently. Since there was no need to explain both phenomena in an integrated way and inevitably admit the presence of an epenthetic segment, different sets of rules are established for each phenomenon in his study. For the tensification phenomenon, he argued that tensification is derived by a tensification rule in which the W_B initial lax onset is simply tensified when the W_A final-segment is sonorous, not by inserting and deleting an epenthetic segment.

Ito (2014) deals only with the compound tensification phenomenon in her study. Thus, there was no need to explain both tensification and nasal lengthening in an integrated way, either. Just as in Chung (1980), she also adopts a way of simply tensifying the W_B initial onset segment. For doing so, she adopts a floating [tense] feature as an underlying representation for the compound juncture marker. In the present study, I also focus on compound tensification. I follow Ito (2014) and assume that a [tense] feature is inserted between two nouns when they form a compound.

Regarding the realization of this [tense] feature, I adopt REALIZEMORPHEME constraint (Kurisu 2001), in accordance with Ito's (2014) analysis. This constraint requires that the floating [tense] feature have phonological exponence. In order to have phonological exponence, the floating feature needs to be linked to an available host segment. First of all, this [tense] feature cannot dock onto the W_A final coda, because codas are unreleased and hence cannot be tensified, according to the general syllable structure constraint of Korean phonology (Kim-Renaud 1974: 104). Thus, the [tense] feature would rather dock onto the W_B initial onset segment in order to be phonologically manifested. In sum, the REALIZEMORPHEME constraint causes the floating [tense] feature to be realized, by the association to the initial onset consonant of W_B.

I also adopt IDENT(tense) which requires that the specification for the [tense] feature of the input segments be identical to that of the output segments. In this study, all the initial onset segments of W_B are underlyingly lax, in other words, without [tense]. Therefore, IDENT(tense) basically blocks these onset segments from bearing a [tense] feature and undergoing tensification. These two constraints responsible for the application of compound tensification are defined in (2).

(2) Constraints responsible for the occurrence of compound tensification

a. REALIZEMORPHEME (RM): Morphemes are phonologically realized (Kurisu 2001).

b. IDENT(tense): Correspondent segments in the input and output are identical for the feature [tense].

Variable occurrences of tensification in compounds come from the relative weights of the constraints in (2). Whereas REALIZEMORPHEME promotes the floating [tense] feature to be phonologically manifested, the faithfulness constraint IDENT(tense) requires the initial onset consonant of W_B to remain unchanged. Depending on the relative importances of these two constraints, whether or not tensification applies to a compound is determined. A sample analysis is shown in (3).

(3) Analysis on the optional tensification of compounds

/san+[tense]+pul/	RM	ID(tns)
i. san.pul	*	
ii. san.p*ul		*

Between the two constraints, if REALIZEMORPHEME outranks the faithfulness constraint IDENT(tense), the candidate (3.ii) wins and tensification occurs. If it is the other way around, the candidate (3.i) wins and tensification does not occur.

4.2. Frequency or Syllable number?

As mentioned in §3.3, some earlier literature proposed that polysyllabic stems tend to resist compound tensification (Lee 1972: 466-467, Kim 1974: 136, Kim-Renaud 1974: 166, and Ahn 1985: 90-91). Zuraw (2011) proved through a corpus study that compound tensification is indeed less probable when W_A and W_B both contain more than one syllable. My data also confirmed this. The tensification rate dropped when the syllable number changed from *one* to *more than one* in both W_A and W_B (see Table 4). Also, it was proven through a Pearson's χ^2 test with Yates' continuity correction that *multisyllabic* is significantly different from *monosyllabic* in its effect on tensification. Thus the syllable number seemed to significantly influence the probability of tensification. Nevertheless, this syllable number effect did not clearly emerge in the mixed effect logistic regression model presented in Table 10. Then, why was the syllable number effect not significant in the regression model?

It is widely accepted and agreed upon that frequent words tend to be short (Zipf 1935: 22). I assume that the syllable number effect was not confirmed in the regression model because this effect was already covered by the frequency factor, which was significant in the regression model as seen in Table 10. Some statistical analyses support this assumption. First of all, a linear regression analysis, with the syllable number as the dependent variable and the frequency as the independent variable, shows a significant result (est. = -0.002, $p < 0.001$). Thus, the general negative correlation between the frequency and the syllable number is also confirmed in my data. Second, excluding either one of these two factors does not significantly drop the accuracy of the regression model, which implies that the syllable number effect and the frequency effect overlap. Using an *ANOVA* function for a likelihood-ratio test in R, I checked that the two mixed effect logistic regression models, one of which includes both W_A/W_B syllable number and frequency as the factors and the other which includes W_A/W_B syllable number but excludes frequency, are not significantly different ($\chi^2 = 11.859$, df = 1, $p < 0.001$, logLik = -2720.1; -2726.0). Furthermore, I also checked that the two models, one of which includes both W_A/W_B syllable number and frequency as the factors and the other which includes frequency but excludes W_A/W_B syllable number, are not significantly different ($\chi^2 = 6.359$, df = 2, $p < 0.05$, logLik = -2720.1; -2723.3). This implies that the syllable number and the frequency are not entirely separate factors. Therefore, it seems a plausible deduction that the syllable number effect was not significant in the mixed effect logistic regression model because this effect was already covered by the frequency effect.

Although these two factors account for nearly the same effect, what actually influences compound tensification is undoubtedly the frequency, not the syllable number. Most of all, the regression model including both these factors showed that the frequency factor is significant while the syllable number factor is not, as seen in Table 10. Moreover, in the previous paragraph, the log likelihood, which is an important measure of the goodness of the model, was higher for the model that included the frequency as a factor, compared to the log likelihood for the model that included the syllable number as a factor (-2723.3 > -2726.0). This indicates that the model including the frequency factor is more accurate than the model including the syllable number factor. In sum, although the effects of these two factors almost overlap, the frequency better explains the data than the syllable number does. That might be the reason why the frequency factor was significant whereas the syllable number was not in the regression model. All things considered, I only establish a constraint for the frequency effect and exclude the one for the syllable number effect from the constraint set, which will be used to capture the pattern of Seoul Korean compound tensification.

Before ending the discussion on the syllable number effect, let me introduce how significant this effect was in Ito's (2014) study on Yanbian Korean. Ito (2014) also confirmed that the syllable number significantly affects the probability of compound tensification. Just like in Zuraw's (2011) study, the greater the syllable number is, the less likely tensification is to apply. However, Ito (2014) claimed that the probability of tensification correlates more linearly with the syllable number of W_A and W_B, than what was reported in Zuraw (2011). More specifically, in Zuraw (2011), the tensification rate dropped substantially when the syllable number changed from monosyllabic to bisyllabic. However, when the syllable number subsequently changed from bisyllabic to trisyllabic, the tensification rate only declined a little. Unlike in Zuraw (2011), in Ito (2014), the tensification rate decreased along with the increasing syllable numbers more steadily and consistently. I speculate that Ito (2014) could observe more linear correlation between the syllable number and the tensification rate because she investigated Yanbian Korean, which has a pitch accent system unlike Seoul Korean. Under the pitch accent system in Yanbian Korean, compound words have only one pitch peak. If a compound has four syllables or more, speakers may separate it into two separate prosodic words and hence may no longer perceive it as a single prosodic unit for an accent assignment. Since this tensification phenomenon is what occurs in a compounding process, the requirement for this process is that the boundary between two nouns be not explicit. If the two nouns are clearly separate, compound tensification would not be applied (Chung 1980: 36). In sum, longer words would undergo tensification less frequently, because these longer words are more likely to be separated and no longer perceived as a single prosodic unit. In Ito's (2014) study, both the frequency, which was instead measured by familiarity ratings, and the syllable number were significant in a single regression model, unlike in my study. If the syllable number effect in Ito's (2014) data indeed came from the pitch system of the Yanbian dialect, it is understandable that the same syllable number effect did not appear in my regression model on Seoul Korean data.

Now, let me start the discussion of the frequency factor. Before establishing a constraint regarding the frequency effect, I first introduce how significant the frequency effect was in the two previous studies on the variation in Korean compound tensification. First, Zuraw (2011) mentioned that compounds that do not undergo tensification all nearly have the lowest frequency, and most of these are not even found in the frequency database because they are not frequent enough to appear in a given corpus. On the other hand, compounds that are known to undergo tensification in the literature mostly have non-zero frequencies. This result implies that compound tensification is more often applied to more frequent lexical items.

Ito (2014) looked at the word-familiarity ratings instead of the frequency, based on the fact that word-familiarity ratings and the frequencies are correlated (Nusbaum *et al.* 1984). In order to determine the word-familiarity ratings for the survey items, she asked four native speakers to judge the familiarity of the words as a numerical value and calculated the averaged familiarity rating for each survey item. With this data, she showed that the averaged familiarity correlated with the tensification rate: for the items with higher familiarity ratings, tensification tends to apply more. As illustrated in §3.2, my data also showed a positive correlation between the frequency and the tensification rate: the higher the frequency is, the more frequently tensification is to apply (see Figure 2). Therefore I adopt a constraint that suppresses tensification for low frequency items.

(4) *TENSE/LOWFREQUENCY (*T/LF): Tensification is blocked for low frequency items (slightly adapted from Ito 2014: 371).

Since the frequency is not a categorical but a linear factor, setting a criterion for the low frequency is required. In this study, the low frequency was defined to be below 33.7, which is an average value of all the frequencies from the survey items. Thus, the constraint in (4) assigns a violation mark to the tensified candidate of a compound whose frequency is below 33.7. A sample analysis is presented in (5).

(5) Analysis of the frequency effect

a. /na.mu + cul.ki/ 'a trunk of a tree' (Frequency: 14)

/ na.mu+[tense]+cul.ki/	RM	ID(tns)	*T/LF
i. na.mu.cul.ki	*		
ii. na.mu.c*ul.ki		*	*

b. /son + pa.tak/ 'palm' (Frequency: 280)

/son+[tense]+pa.tak/	RM	ID(tns)	*T/LF
i. son.pa.tak	*		
ii. son.p*a.tak		*	

Tensification is applied to both candidates (5a.ii) and (5b.ii). However, the candidate (5a.ii) violates *TENSE/LOWFREQUENCY due to its low frequency, whereas the candidate (5b.ii) does not. This is because candidate (5b.ii) has a high frequency and does not meet the condition under which *TENSE/LOWFREQUENCY is active. The violation of this constraint makes tensified candidates of low frequency compounds, such as (5a.ii), less probable, which results in the lower tensification rates for the relatively rarely occurring compounds.

4.3. W_A final-segment type

In §3.4, the tensification rates according to each W_A final-segment type were presented. The tensification rate was the highest when W_A ends with an obstruent coda. The rate decreased as the W_A final-segment type changed from a nasal to a liquid, and decreased again when it changed from a liquid to a vowel (see Table 5). And these four coda types were significantly different from one another in their effects on tensification.

This chapter provides the analysis of each W_A final-segment type. The obstruent coda is discussed in §4.3.1, and the sonorant codas are discussed in §4.3.2.

4.3.1. Post-obstruent tensification in compounds

Speaking of obstruent codas, we must take into account that post-obstruent tensification (henceforth POT), in which a lax onset consonant gets tensified after an obstruent coda, is known to be obligatory in Korean. Since a lax consonant in this context always undergoes tensification even when the tensification is not triggered by a compounding process, there is no way to tell the compound tensification and POT apart (Moon 1997, Labrune 1999: 135). Table 11 shows the two derivations of the same compound, one with and the other without the compound juncture marker, the epenthetic /t/.

UR	/pap + kɨ.rɨt/	/pap + kɨ.rɨt/
Epenthetic /t/	/pap + **t** + kɨ.rɨt/	N/A
Post obstruent tensification	/pap + t + k*ɨ.rɨt/	/pap + k*ɨ.rɨt/
Cluster simplification	/pap + k*ɨ.rɨt/	N/A
SF	[pap.k*ɨ.rɨt]	[pap.k*ɨ.rɨt]

Table 11

The two derivations with and without the insertion of the epenthetic /t/. Whether or not the epenthesis takes place, both derivations result in the identical surface form.

When the epenthesis takes place, the tensified surface form is derived through the automatic rules of Korean as shown in Table 11. Even when the epenthesis does not occur, the final coda of W_A, the obstruent /p/, tensifies the immediately following onset consonant in W_B, resulting in the identical surface form. In sum, regardless of whether or not tensification is triggered by the compound juncture marker, both surface forms are equally tensified through POT.

For this reason, Zuraw (2011) excluded compounds with obstruent-final W_A from her analysis. She mentioned that the insertion of *saisiot* shows no effect in compound tensification when the W_A final coda is an obstruent. More importantly, her analysis was based on a Korean dictionary, where pronunciations reflect automatic phonological rules of Korean, including POT. Therefore in her data, the pronunciation of all underlyingly lax consonants after an obstruent coda must have been marked as tense, resulting in the 100% tensification rate after an obstruent coda.

By contrast, Ito (2014) included compounds with obstruent-final W_A in her analysis. Since she collected the large-scale pronunciation data from native Korean speakers, not from a dictionary, she could check if compound tensification applies variably even after an obstruent coda. Base on the data collected through her own survey, she reported that the tensification rate after an obstruent coda was 69%, which might imply that POT does not automatically apply in the Yanbian dialect.

Either in my survey, the tensification rate after an obstruent coda did not reach 100%, which might indicate that POT is not obligatorily applied at the juncture of a compound word. Jun (1993, 1998) and Shin (1999) showed by conducting phonetic experiments that the domain of POT is an accentual phrase (henceforth AP). That is, POT is applied within a single AP, but not across an AP boundary. For example, it was shown in Jun (1993: 110-112, 1998: 209) that a verb phrase /mi.yək.kuk + pə.ryə/ 'throw out the seaweed soup' undergoes POT if this phrase constitutes a single AP as in $_{AP}$[mi.yək.k*uk.**p***ə.ryə], but does not undergo this same phonological process if this phrase is produced as the two parsed accentual phrases as in $_{AP}$[mi.yək.k*uk] $_{AP}$[**p**ə.ryə]. Similarly, in Shin (1999: 34), it was shown that a noun phrase /ma.ci.mak + ka.ɨl/ 'the last autumn' undergoes POT in $_{AP}$[ma.ci.mak.**k***a.ɨl], but does not undergo POT in $_{AP}$[ma.ci.mak] $_{AP}$[**k**a.ɨl]. In sum, these two studies claimed that the application of POT depends on whether or not the trigger and the target of POT are in a single AP. If both are in the same AP, the target lax obstruent becomes tense, but when there is an AP boundary between the target and the trigger, the target lax obstruent remains unchanged.

I speculate that the tensification rate with W_A obstruent-final words did not reach 100% in my survey because there was often an AP boundary between the W_A final obstruent coda and the W_B initial lax onset. In other words, W_A and W_B were often produced as the two independent APs. When this happened, the application of POT might have failed and the pronunciation form without tensifying W_B initial onset consonant might have been chosen by the participants.

Despite not reaching 100%, the tensification rate after an obstruent coda is still higher than the rates after the other coda types, which might indicate that POT still plays a significant role at the compound juncture. In order to capture the highest tensification rate after an obstruent coda, I adopt a constraint that blocks a sequence of an obstruent and a lax consonant in a single AP, as shown in (6). The tableaux in (7) illustrate the sample analyses.

(6) *obs-lax: No lax obstruent after an obstruent in a single AP.

(7) Analysis of POT

a. /ap^h + ta.ri/ ‘a front leg’

/ap^h+[tense]+ta.ri/	*obs-lax	RM	ID(tns)
i. ap.ta.ri	*	*	
ii. ap.t*a.ri			*

b. /tol + tam/ ‘a stone wall’

/tol+[tense]+tam/	*obs-lax	RM	ID(tns)
i. tol.tam		*	
ii. tol.t*am			*

Tableaux (7a) and (7b) illustrate an analysis of a case in which W_A ends with an obstruent coda and with a sonorant coda, respectively. Between the two candidates in (7a), the candidate without tensification (7a.i) violates *obs-lax, whereas the candidate with tensification (7a.ii) does not. Unlike in (7a), both candidates in (7b) vacuously satisfy *obs-lax, since the W_A final coda is not an obstruent in (7b). The violation of *obs-lax makes (7a.i) less probable than (7b.i), which results in the higher probability of tensification after an obstruent coda compared to those after sonorants.

4.3.2. Lenition versus tensification

Regarding the sonorous W_A final-segment types, I first introduce the results of the two previous studies and compare those with my results. And then, I provide the analysis of the W_A final sonorant codas, based on the notion of lenition as opposed to tensification, which is a kind of fortition.

Zuraw (2011) reported that the propensity for tensification is the highest when W_A ends with a liquid, lower with a nasal, and the lowest with a vowel. Ito (2014) reported that the rate of tensification in Yanbian Korean is higher when W_A ends with a sonorant coda, which refers to both liquid and nasal codas, compared to when W_A ends with a vowel. In my survey, the likelihood for tensification decreased in post-nasal, post-liquid and post-vowel environments, in that order.

My study agrees with the two previous studies in that the tensification rate is the lowest after a vowel, but disagrees in whether the effect on tensification is significantly different between nasal and liquid W_A final codas. In the two studies prior to this one, nasal and liquid codas did not make a significant difference in their effects on tensification. First, Zuraw's (2011) regression model yielded a large p-value on the change from nasal to liquid coda (est. = 0.06, p = 0.625)[4], which indicates that these two codas do not make a significant difference in tensification. Also, in Ito (2014), no clear difference was observed between post-liquid and post-nasal environments (tensification rate: *post-nasal* 59%, *post-liquid* 57%). She therefore grouped these two categories into a single category, *sonorant*. In my data, the tensification rates were quite different between W_A nasal-final compounds and W_A liquid-final compounds: tensification is more likely when W_A ends with a nasal, compared to when it ends with a liquid.

Meanwhile, Ito (2014) suggested that the impetus for lenition might militate against the application of tensification. And different preferences on lenition in various contexts would lead to different tensification rates among the W_A final-segment types. More specifically, if a given context is a preferred target of lenition, a consonant would rather undergo lenition, not tensification. Since a consonant in an intervocalic position is a widely attested target of lenition, Ito (2014) assumes that tensification rate is the lowest when W_A ends with a vowel.

Kirchner (1998) gives an insightful analysis on lenition, based on an effort-based approach. He argues that the motivation for lenition, whereby a sound becomes weaker, is basically to minimize the articulatory effort. The articulatory effort cost is measured by the displacement of an articulator involved, mainly the

---

[4] Zuraw's (2011) logistic regression model:

	Est.	SE	z	P (>|z|)
W_A ends with,				
nasal (vs. V)	0.49	0.07	6.19	< 0.001 ***
liquid (vs. nasal)	**0.06**	0.12	0.48	**0.625**
obstruent (vs. liquid)	0.15	0.23	0.66	0.509

Although it shows that the probability of compound tensification becomes higher in the post-liquid position compared to the post-nasal position (est. = 0.06), the difference between these two contexts is not significant (p = 0.625).

jaw. For example, when making vowel sounds, the jaw position is relatively lower compared to when pronouncing consonants. Therefore, the displacement of the jaw is greater and hence, more effort is required when pronouncing a sequence of a vowel and a consonant, compared to a sequence of two consonants, because a change from a vowel to a consonant requires longer movement of the jaw. Since the purpose of lenition is the reduction of such articulatory efforts, depending on the effort required for making the gesture, lenition occurs with different probabilities. The more effort a gesture requires, the more likely lenition is to apply. For example, a consonant is more prone to undergo lenition when it is adjacent to a vowel, compared to when it is adjacent to another consonant, due to the higher degree of effort or displacement required for the former context. On this basis, he suggested the lenition-trigger hierarchy, as seen in (8).

(8) Lenition-trigger hierarchy (Kirchner 1998: 197)

Vowels > **liquids** > glides > **nasals** > **stops** > strident fricatives > geminate

This hierarchy in (8) reflects the degree of the preference for lenition in each context. As mentioned above, this study pointed out that the effects of nasal and liquid codas on tensification are significantly different from each other. According to the lenition-trigger hierarchy, the impetus for lenition is higher when a target consonant is adjacent to a liquid than it is adjacent to a nasal. These different impetuses for lenition between the two segments accord with the different tensification rates between post-nasal and post-liquid environments. Thus, the result that tensification was applied less after a liquid coda than after a nasal coda is corroborated by this lenition-trigger hierarchy.

In this study, following Ito (2014), it is also assumed that these different degrees of motivation for lenition bring about the different tensification rates after each preceding segment. The different constraints are responsible for blocking tensification in the different contexts. Therefore, I establish a constraint for each preceding segment type, as shown in (9).

(9) Constraints for the sonorous W_A final-segment types (slightly adapted from Ito 2014: 371).

a. *TENSE/VOWEL__ (*TNS/V_): After a vowel, no tensification.

b. *TENSE/LIQUID__ (*TNS/L_): After a liquid, no tensification.

c. *TENSE/NASAL__ (*TNS/N_): After a nasal, no tensification.

According to the lenition-trigger hierarchy, the impetus for the lenition process is the greatest when W_A ends with a vowel, lesser with a nasal, and the least with a liquid. If the impetus for lenition is greater, the tensification process, which is usually considered as a type of fortition, is more strongly blocked. Thus other things being equal, it is expected that the weights of these three constraints, which block tensification, are as follows: *TENSE/VOWEL-W_A > *TENSE/LIQUID-W_A > *TENSE/NASAL-W_A. The analyses of each W_A final-segment type are shown in (10).

(10) Analysis on the sonorous W_A final-segment types

a. /kæ.mi + cip/ 'an ants' nest'

/kæ.mi+[tense]+cip/	RM	ID(tns)	*TNS/V_	*TNS/L_	*TNS/N_
i. kæ.mi.cip	*				
ii. kæ.mi.c*ip		*	*		

b. /tol + tam/ 'a stone wall'

/tol+[tense]+tam/	RM	ID(tns)	*TNS/V_	*TNS/L_	*TNS/N_
i. tol.tam	*				
ii. tol.t*am		*		*	

c. /nun + sa.ram/ 'a snowman'

/nun+[tense]+sa.ram/	RM	ID(tns)	*TNS/V_	*TNS/L_	*TNS/N_
i. nun.sa.ram	*				
ii. nun.s*a.ram		*			*

In tableaux in (10), the three candidates (10a.ii), (10b.ii), and (10c.ii), violate the corresponding constraint in (9), for tensifying their W_B initial onset segment after the corresponding coda. However, the motivation for blocking tensification is the greatest when W_A ends with a vowel and thus presumably *TENSE/VOWEL-W_A would be assigned the greatest numerical weight among the three constraints in (9). By the gradiently assigned weights of these three different constraints, which are responsible for blocking tensification in each specific context, the different tensification rates among the different coda types are to be captured. The greater weight of the constraint *TENSE/VOWEL-W_A will make the candidate (10a.ii) less probable than (10b.i) and (10c.i) in order, which leads to tensification rates increasing in the same order.

4.4. *Leakage*

In §3.5, it was confirmed from the survey results that when W_A ends with a liquid, coronal consonants (*s*, *c* and *t*) in W_B initial onset position undergo tensification much more readily than non-coronal consonants (*k* and *p*). Such an asymmetry between the coronal and the non-coronal consonants seems to mirror the pattern of the existing lexicon. In the native Korean simplex word lexicon, the frequency of the tense coronal obstruents (*s**, *c** and *t**) after coda *l* is several times higher than that of the tense non-coronal ones (*p** and *k**) (Ko 1996: 40). This is because in Middle Korean, after a coda *l*, the immediately following coronal lax obstruent underwent tensification, while non-coronal ones remained intact.

Therefore, in Contemporary Korean, while sequences like *ls**, *lt**, and *lc** are easily found as in /kɨl.**s***i/ 'handwriting', /p^hal.**t***uk/ 'forearm', and /nal.**c***a/ 'date', their lax counterparts *ls*, *lt*, and *lc* are not found in monomorphemic words (Aum 1998: 10). Actually, this pattern has been formulated as a morpheme structure rule, which by its definition holds within a single morpheme: /s, t, c/ -> [s*, t*, c*] / *l* $ __ (Ko 1996)[5]. Although this rule is active in the tautomorphemic domain, which is the reason why it is called a morpheme structure rule, I hypothesize that its effect diffuses to the heteromorphemic domain as well.

[5] '$' stands for a syllable boundary.

Martin (2011) supports this hypothesis. In his study, it was proven that a phonotactic constraint, which holds within morphemes, diffuses its weaker effect across morpheme boundaries. In other words, although violating a morpheme-internal phonotactic constraint is completely licensed in compounds, this constraint is violated in compounds only up to a certain point. This process was named *leakage*, in the sense that a phonotactic generalization somewhat leaks from the tautomorphemic domain to the heteromorphemic domain.

Applying this concept to the case of compound tensification, it can be stated that the weaker version of the morpheme-internal phonotactic constraint, which bans the sequence of a liquid coda and a lax coronal onset, holds across morpheme boundaries. Thus, although it is licensed, a sequence of a liquid coda and a lax coronal onset is less preferred than that of a liquid coda and a tense coronal onset at the morpheme boundary of a compound. Therefore, a coronal onset after a liquid coda tends to undergo compound tensification more frequently, reflecting the asymmetric pattern that exists within a single morpheme.

In addition to native Korean compound words, this *leakage* phenomenon is seen also in Sino-Korean words. In Sino-Korean words, coronal lax onsets following a liquid coda frequently undergo tensification as in /hwal + toŋ/ [hwal.t*oŋ] 'activity', /pal + sæŋ/ [pal.s*æŋ] 'occurrence', and /səl.cəŋ/ [səl.c*əŋ] 'establishment' (Kim-Renaud 1974: 171-172, Ko 1996: 41, and Byeon 2012: 233). Although this tensification process in Sino-Korean words is known to be automatic, a number of the previous studies argued or proved that this tensification process in Sino-Korean words is just a general tendency with many exceptions (Lee 1996, Aum 1998: 12, Kim 2003: 19-21, Ha 2006: 35-36, and Lee 2009: 156). For example, Lee (1996, 2009) showed that the tensification is optionally applied in

/mil + sil/ [mil.sil] ~ [mil.s*il] 'a secret room', /il + sik/ [il.sik] ~ [il.s*ik] 'Japanese food', and /cəl + sak/ [cəl.sak] ~ [cəl.s*ak] 'cutting', and is not applied in /t^hɨk.pyəl + si/ [t^hɨk.pyəl.si] 'metropolitan city', /cæŋ.t^hal + cən/ [cæŋ.t^hal.cən] 'fight', and /pyəl + to.ri/ [pyəl.to.ri] 'an alternative'. I speculate that this tensification tendency in Sino-Korean words is also the weaker version of the morpheme-internal phonotactic constraint. Considering that each syllable in Sino-Korean words corresponds to a single morpheme, this can be additional evidence for grammar *leakage* from the tautomorphemic domain towards heteromorphemic domain.

Meanwhile, it is assumed in Martin (2011) that *leakage* is a result of the learning mechanism. Language learners are first exposed to the input data, in which monomorphemes violating a given phonotactic constraint are absent. To the learners equipped with this biased data, compounds violating the constraint would sound worse, which will cause the learners to prefer compounds that obey the constraint. On this assumption, both tautomorphemic and heteromorphemic phonotactics, despite the difference in strength, result from the same constraint. To capture this difference in their strengths, Martin (2011) establishes two different constraints. One is called a structure-sensitive constraint, which takes into account the morphological structure and only penalizes illegal forms that exist within a specific domain. The other one is called a structure-blind constraint, which ignores the morphological structure and hence penalizes illegal forms within and across a specific domain.

It is already certain that a sequence of a liquid and a lax coronal obstruent is not found in Korean monomorphemic words, which indicates that a structure-sensitive constraint is always obeyed. Thus, adopting a structure-sensitive constraint, which is for the monomorphemic words, is somewhat unnecessary. What is examined here is whether or not the effect of the morpheme-internal phonotactic constraint emerges across morpheme boundaries. Therefore, only the structure-blind constraint is included in the analysis of Seoul Korean compound tensification. I adopt a broad constraint that penalizes a sequence of liquid coda and lax coronal onset, either within or across the morpheme boundaries. Although it ignores morphological structure, it will only penalize the sequence of a liquid coda and a lax coronal onset consonant which exists across morpheme boundaries, since there is no case in which this constraint is violated within a single morpheme anyway.

(11) *L(+)C: A sequence of a liquid coda and a lax coronal onset is not allowed[6] (borrowed the constraint format from Martin 2011: 761).

As mentioned in §4.1, this study assumes that REALIZEMORPHEME brings out the realization of [tense] feature and triggers compound tensification in the first place. By adding the constraint *L(+)C, the different preferences for tensification between the coronal and the non-coronal consonants after a liquid coda will be captured. The tableaux in (12) illustrate an analysis of *leakage*.

(12) Analysis of *leakage*

a. /tol + sot^h/ 'a stone pot'

/tol+[tense]+sot^h/	RM	ID(tns)	*L(+)C
i. tol.sot	*		*
ii. tol.s*ot		*	

[6] '(+)' means that the morpheme boundary '+' is optional. Thus this constraint *L(+)C penalizes a sequence of liquid coda and coronal lax onset consonant within and across morphemes.

b. /pal + kup/ ‘a hoof’

/pal+[tense]+kup/	RM	ID(tns)	*L(+)C
i. pal.kup	*		
ii. pal.k*up		*	

Candidate (12a.i) violates not only the tensification triggering constraint REALIZEMORPHEME but also the constraint *L(+)C, while candidate (12a.ii) only violates IDENT(tense). This makes the candidate containing the tensified coronal consonant (12a.ii) more probable than (12a.i). By contrast, both candidates in (12b) are not influenced by *L(+)C at all because the place of articulation of W_A initial onset consonant is velar. The additional violation of *L(+)C by (12a.i) guarantees the relatively higher occurrence of (12a.ii) compared to (12b.ii), which might explain the asymmetric behavior between the coronal and the non-coronal consonants after a liquid.

4.5. OCP effect

Zuraw (2011) showed that having a tense consonant in either W_A or W_B prevents tensification in Seoul Korean, with a stronger effect of W_A. In Ito (2014), the many aspects of the OCP effect in compound tensification are extensively examined. She showed that not only a tense consonant but also an aspirated one can block tensification in Yanbian Korean, with no significant difference between the two.

Also, she showed that compared to W_A, the effect is stronger when either a tense or aspirated consonant exists in W_B. Moreover, the results of her wug test on nonce words showed that an aspirated consonant in W_A tends to resist tensification more strongly when it is in a syllable adjacent to the site of tensification. This locality effect was unclear in existing words but emerged in nonce words.

In my data, as it was reported in §3.6.1, while a tense or an aspirated consonant in W_B clearly showed the OCP effect, either one in W_A did not show any OCP effect. This is similar to Japanese Rendaku in that the presence of a voiced obstruent in W_B blocks voicing of the W_B initial consonant, while the presence of one in W_A does not show a substantial blockage effect (Ito & Mester 1998: 29, Kawahara & Sano 2014: 115-118). From the case of Seoul Korean compound tensification and Japanese Rendaku, it can be stated that at least for these two phenomena the blocking effect on marking the compound juncture comes from the OCP effect within a stem. In other words, the OCP constraint that bans a co-occurrence of laryngeally marked segments or voiced obstruents displays only within a stem domain.

In this study on Seoul Korean compound tensification, capturing this OCP effect exerted by the presence of a laryngeally marked consonant in W_B, I adopt a constraint that blocks tensification if an aspirated or a tense consonant is contained in the same stem as the site of tensification, namely in W_B. The constraint is illustrated in (13). Sample analyses are shown in (14).

(13) OCP(stem): Do not allow a co-occurrence of laryngeally marked consonants in a stem (slightly adapted from Coetzee 2004: 364).

(14) Analysis of the OCP effect

a. /k^hoŋ + ki.rɨm/ 'soybean oil'

/k^hoŋ+[tense]+ki.rɨm/	RM	OCP(stem)	ID(tns)
i. k^hoŋ.ki.rɨm	*		
ii. k^hoŋ.k*i.rɨm			*

b. /mal + ta.t^hum/ 'quarrel'

/mal+[tense]+ta.t^hum/	RM	OCP(stem)	ID(tns)
i. mal.ta.t^hum	*		
ii. mal.t*a.t^hum		*	*

The constraint OCP(stem) requires that laryngeally marked consonants do not co-occur in a single stem. In other words, this constraint will penalize a compound that contains a laryngeally marked consonant in its W_B word and also undergoes tensification. In tableaux (14), the candidate (14a.ii) contains an aspirated consonant /k^h/ in W_A, while the candidate (14b.ii) contains an aspirated one /t^h/ in W_B. Therefore, the constraint OCP(stem) will penalize only (14b.ii) for undergoing compound tensification with the presence of an aspirated consonant /t^h/ in W_B. The violation of OCP(stem) by (14b.ii) will lead to the lower occurrence of (14b.ii) than that of (14a.ii), capturing the OCP effect that displays within the stem domain.

4.6. Summary

So far in this section, not only the constraints responsible for the occurrence of compound tensification but also those responsible for the various phonological trends observed in my survey results were established. All the constraints employed for the analysis of Seoul Korean compound tensification are summarized in (15).

(15) Constraint set adopted for the analysis of Seoul Korean compound tensification

a. REALIZEMORPHEME (RM): Morphemes are phonologically realized.

b. IDENT(tense): Correspondent segments in the input and output are identical for the feature [tense].

c. *TENSE/LOWFREQUENCY (*T/LF): Tensification is blocked for low frequency items.

d. *obs-lax: No lax obstruent after an obstruent in a single AP.

e. *TENSE/VOWEL-W_A (*TNS/V_): If W_A ends with a vowel, no tensification.

f. *TENSE/LIQUID-W_A (*TNS/L_): If W_A ends with a liquid, no tensification.

g. *TENSE/NASAL-W_A (*TNS/N_): If W_A ends with a nasal, no tensification.

h. *L(+)C: A sequence of liquid coda and coronal lax onset is not allowed.

i. OCP(stem): Do not allow a co-occurrence of laryngeally marked segments in a stem.

5. Learning simulation

In the previous section, the constraint set for the analysis of Seoul Korean compound tensification was established. In this section, I will conduct a learning simulation using the computer software Maxent Grammar Tool (Hayes 2009a). By doing so, I will show not only that this grammar can be learned from the actual data but also that the learned grammar can capture the observed patterns of Seoul Korean compound tensification.

I will first try to find the specific weight of each constraint in (15), by conducting a learning simulation on the training data which are basically the results of my own survey on compound tensification. Then, testing whether this weighted constraint set successfully reproduces the observed distribution of the candidates, I will apply these constraints to a test set of words which is essentially identical to the training data.

The training data consist of the eighteen word types, which are classified according to the several criteria that reflect the trends observed in the survey results. Thus it is expected that each word type would behave differently toward the various constraints in (15). The two possible candidates, one with and the other without tensifying the initial onset consonant of W_B, are presented for each compound word type. Then, these two candidates are provided with their frequencies, which are taken from the relevant distribution in the survey results. The training data is as summarized in Table 12.

W_A coda	W_B ons plc.	Frequ-ency	OCP W_B	Response type	Frequency	Proportion
Vowel		Low	O	Tensified	26	0.177
				Non-tensified	121	0.823
			X	Tensified	474	0.434
				Non-tensified	618	0.566
		High	O	Tensified	1	0.016
				Non-tensified	62	0.984
			X	Tensified	150	0.476
				Non-tensified	165	0.524
Liquid	Coronal	Low	O	Tensified	22	0.349
				Non-tensified	41	0.651
			X	Tensified	392	0.566
				Non-tensified	301	0.434
		High	O	Tensified	11	0.131
				Non-tensified	73	0.869
			X	Tensified	146	0.772
				Non-tensified	43	0.228
	Non-coronal	Low	O	Tensified	0	0.000
				Non-tensified	21	1.000
			X	Tensified	324	0.406
				Non-tensified	474	0.594
		High	X	Tensified	52	0.413
				Non-tensified	74	0.587
Nasal		Low	O	Tensified	7	0.083
				Non-tensified	77	0.917
			X	Tensified	771	0.633
				Non-tensified	447	0.367
		High	X	Tensified	454	0.832
				Non-tensified	92	0.168
Obstruent		Low	O	Tensified	117	0.796
				Non-tensified	30	0.204
			X	Tensified	461	0.878
				Non-tensified	64	0.122
		High	O	Tensified	13	0.619
				Non-tensified	8	0.381
			X	Tensified	223	0.885
				Non-tensified	29	0.115

Table 12

Training data given to Maxent Grammar Tool for assigning weights to the set of constraints.

In the training data, the two combinations could not be included, due to the absence of compounds with such combinations. These two combinations are as follows: (a) liquid-final W_A, non-coronal W_B onset, high frequency, and a laryngeally marked consonant in W_B, and (b) nasal-final W_A, high frequency, and a laryngeally marked consonant in W_B.

The training data shown in Table 12 are input to Maxent Grammar Tool, with a default setting (mu = 0; σ^2 = 100,000). The weights obtained are shown in Table 13.

OCP(stem)	1.6766306
*obs-lax	1.5335025
REALIZEMORPHEME	1.2273501
*TENSE/LIQUID-W_A	1.1021664
*TENSE/VOWEL-W_A	0.9832487
*L(+)C	0.7857380
*TENSE/LOWFREQUENCY	0.4292537
IDENT(tense)	0.1427664
*TENSE/NASAL-W_A	0.0000006

Table 13

Weights obtained using Maxent Grammar Tool.

The constraint OCP(stem) has the highest weight, which might reflect the survey results in which the presence of a laryngeally marked consonant in W_B substantially and consistently suppressed tensification.

The tensification triggering constraint REALIZEMORPHEME considerably outranks the faithfulness constraint IDENT(tense). Thus, even when tensification is blocked in a compound, this effect comes from the other markedness constraints, not from the faithfulness constraint.

The four constraints *TENSE/VOWEL-W_A, *TENSE/LIQUID-W_A, *TENSE/NASAL-W_A, and *obs-lax are relevant to W_A final-segment type. Among these, *obs-lax was weighted higher than the other three constraints, not only indicating that POT plays a significant role at the juncture of Seoul Korean compound words, but also mirroring the survey results in which the tensification rate was the highest with the W_A obstruent-final compounds. With regard to the other W_A final-segment types, the constraints *TENSE/VOWEL-W_A and *TENSE/LIQUID-W_A both block tensification when W_A ends with a vowel and a liquid, respectively. Considering the survey results in which the tensification rate was lower with W_A vowel-final compounds than with W_A liquid-final compounds, it was expected that *TENSE/VOWEL-W_A would outrank *TENSE/LIQUID-W_A. However, the result was the opposite: *TENSE/LIQUID-W_A slightly outranks *TENSE/VOWEL-W_A. This result is due to the constraint *L(+)C which is also relevant to the W_A liquid-final compounds. Let me explain with the example tableaux in (16).

(16) Analysis on W_A liquid-final and W_A vowel-final compound

a. /kæ.mi + cip/ 'an ants' nest'

/kæ.mi+[tense]+cip/	*TNS/L_	*TNS/V_	*L(+)C
i. kæ.mi.cip			
ii. kæ.mi.c*ip		*	

b. /tol + sot^h/ ‘a stone pot’

/tol+[tense]+sot^h/	*TNS/L_	*TNS/V_	*L(+)C
i. tol.sot^h			*
ii. tol.$s{*}ot^h$	*		

Tableaux (16a) and (16b) show a sample analysis on the W_A vowel-final compounds and the W_A liquid-final ones, respectively. In (16a), *TENSE/VOWEL-W_A suppresses tensification by penalizing the candidate (16a.ii) and there is no opposing constraint which encourages tensification. By contrast, in (16b), *TENSE/LIQUID-W_A suppresses tensification by penalizing (16b.ii) and there is the opposing constraint *L(+)C which rather promotes tensification by penalizing (16b.i). Although the tensification rate was lower for W_A vowel-final items and hence, it was expected that *TENSE/VOWEL-W_A would outrank *TENSE/LIQUID-W_A, the result of the learning simulation showed the opposite because the tensification blocking effect of *TENSE/LIQUID-W_A is largely offset by *L(+)C. Considering the intervening effect of *L(+)C, the higher weight of *TENSE/LIQUID-W_A is understandable.

Meanwhile, *TENSE/NASAL-W_A weighed almost nothing. The effect of *TENSE/NASAL-W_A is to block tensification when W_A ends with a nasal. Considering that W_A nasal-final compounds had the highest tensification rate among the three sonorous W_A final-segment types, it seems a plausible result that *TENSE/NASAL-W_A weighed about 0.

The role of *TENSE/LOWFREQUENCY, which is to suppress tensification for low frequency items, is also confirmed.

Let me now check if this weighted constraint set can successfully reproduce the observed distribution of the candidates. Table 14 and Figure 4 show the observed and predicted tensification rate of each compound type.

W_A coda	W_B ons plc.	Frequency	W_B OCP	Observed	Predicted
Vowel		Low	O	0.177	0.119
			X	0.434	0.419
		High	O	0.016	0.171
			X	0.476	0.525
Liquid	Coronal	Low	O	0.349	0.208
			X	0.566	0.584
		High	O	0.131	0.287
			X	0.772	0.683
	Non-coronal	Low	O	0.000	0.107
			X	0.406	0.390
		High	X	0.413	0.496
Nasal		Low	O	0.083	0.265
			X	0.633	0.658
		High	X	0.832	0.747
Obstruent		Low	O	0.796	0.625
			X	0.878	0.899
		High	O	0.619	0.719
			X	0.885	0.932

Table 14

The observed and predicted tensification rate of each compound type.

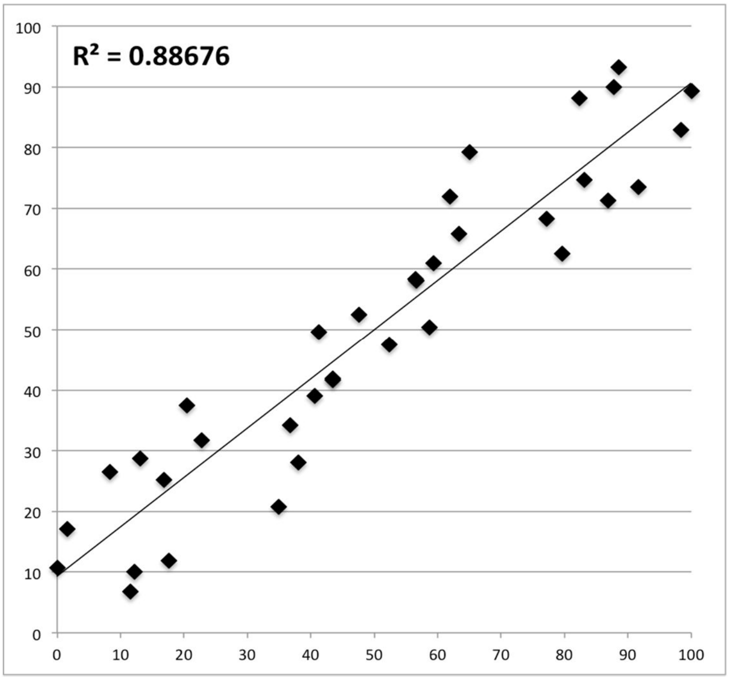

Figure 4

Correlation between the observed and predicted distributions of compound tensification.

At a glance, it can be seen in Table 14 that the posited grammar was highly successful in predicting the tensification rate of each compound type, indicating that the weighted constraint set properly reflects the crucial effects observed in the survey results. A linear regression analysis with the observed distribution as the dependent variable and the predicted distribution as the independent variable shows a significant result ($R^2 = 0.88$, est. = 1.09, SE = 0.06, $p < 2e\text{-}16$).

6. Remaining problems

So far, throughout this study, various tendencies of Seoul Korean compound tensification are introduced. Also, a formal analysis on this phenomenon was developed, in the frame of MaxEnt OT. By doing so, the goals of this study are achieved. This section discusses what problems still remain.

In this study, I performed a survey when collecting the data for investigating the pattern of the compound tensification phenomenon. Among the survey items, there are some W_A vowel-final compounds in which the orthographic regulation requires that *saisiot* (<s>) be specified in the coda position of the W_A final syllable (e.g. /coŋ.i + co.kak/ 'a piece of paper' *<coŋ.i.co.kak>, < coŋ.i**s**.co.kak>). When performing the survey, based on the assumption that the participants would apply more tensification if the coda <s> is manifested, I presented the two components of each compound in separate parentheses in the questionnaire form (see Table 3). Although I did not specify the coda <s> for the compounds in which the orthographic regulations require it to be marked, one might still claim that the participants already bear the prescribed spelling in mind and apply more tensification to these compounds. In other words, the coda <s>, despite being invisible in the questionnaire, could have affected the tensification rates of some W_A vowel-final compounds.

In order to examine the effect of the orthography, I compared the tensification rate of the words with the letter /s/ at the end of W_A and those without it, as shown in Table 15.

Orthographic regulations	Tensification rate (%)
Saisiot (27)	77.4 (439 / 567)
No *saisiot* (50)	20.1 (212 / 1050)

Table 15

Tensification rate (%) as a function of the orthographic requirements for specifying *saisiot*.

Table 15 shows the tensification rate according to whether or not the orthographic regulation requires *saisiot* be specified. The table above shows that tensification is more likely if the orthographic regulation requires that *saisiot* be marked in a compound (χ^2 = 195.73, $p < 2.2e\text{-}16$). These significantly different tensification rates between the two categories indicate that the participants indeed bear the prescribed spelling of the compounds in their minds and react sensitively to the presence of the coda <s>, although it was not manifested in the survey form.

In order to examine the relative significance of the orthography factor, I constructed another mixed effect logistic regression model that basically includes all the significant factors from the model in Table 10, and additionally includes the orthography factor as an independent variable (categorical: *no-saisiot* (ref), *saisiot*). The result is shown in Table 16.

	Estimate	Std. Error	Z value	P(>|z|)
(Intercept)	- 2.016	0.466	- 4.32	< 0.001 ***
Frequency	0.863	0.236	3.65	< 0.001 ***
Orthography (*saisiot*)	3.813	0.551	6.91	< 0.001 ***
W_A coda (liquid-vowel)	0.962	0.554	1.73	0.082
W_A coda (nasal-liquid)	2.155	0.470	4.58	< 0.001 ***
W_A coda (obs-nasal)	1.960	0.692	2.83	0.004 **
W_B onset place (cor)	- 0.106	0.279	- 0.38	0.703
W_A consonant (tense-lax)	0.138	0.377	0.36	0.713
W_A consonant (asp-tense)	- 0.059	0.551	- 0.10	0.913
W_B consonant (tense-lax)	- 2.597	0.583	- 4.45	< 0.001 ***
W_B consonant (asp-tense)	- 0.263	0.873	- 0.30	0.762
W_A coda (L-V): W_B onset (cor)	1.173	0.689	1.70	0.088
W_A coda (N-L): W_B onset (cor)	- 1.584	0.646	- 2.45	0.014 *
W_A coda (P-N): W_B onset (cor)	- 0.672	0.861	- 0.78	0.435

Table 16

Results of a mixed effect logistic regression model including the orthography factor.

The result shows that the orthography factor is indeed significant even when it is incorporated in the model with the other significant factors. The result also shows that compound tensification is more likely to apply if the orthographic regulation requires that *saisiot* (<s>) be specified in the coda position of the W_A final syllable. However, what is important is that the inclusion of this orthography factor as an additional independent variable does not obscure the other significant factors at all. In other words, regardless of whether or not the orthography factor is included, the general tendencies observed in compound tensification still remain robust (compare Table 10 and Table 16). Therefore, it seems a plausible deduction that the main tendencies confirmed throughout this study are quite independent of the orthographic regulation. To conclude, although the participants are affected by the orthography, the general tendencies are still meaningful and significant.

7. Conclusions

In this study, I have investigated compound tensification in Seoul Korean. This tensification phenomenon occurs in a compounding process with various probabilities across compound words. Based on the data collected by performing a survey on twenty-one native Seoul Korean speakers, I confirmed that various phonological and non-phonological factors play a significant role in Seoul Korean compound tensification, and also that these factors contribute to the overall probability of the rule application. The tendencies caused by these factors are summarized as follows: (i) tensification is more likely with high frequency items; (ii) tensification is more likely when W_A (the first component of a compound) ends with an obstruent, followed by a nasal, liquid and a vowel in order; (iii) tensification is less likely when W_B (the second component of a compound) contains a laryngeally marked consonant; (iv) tensification is more likely when W_B ends with a liquid and W_A also begins with a coronal consonant.

In addition, I provided a formal analysis of Seoul Korean compound tensification. The tendencies mentioned above were formalized into the separate OT constraints. Since this phenomenon shows the variable pattern, MaxEnt OT was employed. The specific weights of each constraint are evaluated through Maxent Grammar Tool, based on the input data which is basically the survey results of mine. This weighted set of the constraints was highly successful in predicting the applicability of compound tensification because the reproduced data highly accorded with the distribution of the input data. Thus, it is established that the grammar suggested for Seoul Korean compound tensification can be learned from the actual native speakers' pronunciation data, and also that this learned grammar can capture the various tendencies observed in this phenomenon.

Appendix: the list of Korean compound words

Compound	W_A	W_B	Glossary
a.c^{h}im.pap	a.c^{h}im	pap	'breakfast'
a.ræ.sa.ram	a.ræ	sa.ram	'one's junior'
am.kə.mi	am	kə.mi	'a female spider'
am.ko.yaŋ.i	am	ko.yaŋ.i	'a female cat'
an.kam	an	kam	'lining'
an.sal.rim	an	sal.rim	'housekeeping'
an.c*aŋ.ta.ri	an.c*aŋ	ta.ri	'knock-knee'
an.kæ.pi	an.kæ	pi	'smir'
ap^h.pa.k^{h}wi	ap^h	pa.k^{h}wi	'front wheels'
ap^h.pa.ta	ap^h	pa.ta	'costal waters'
ap^h.ta.ri	ap^h	ta.ri	'foreleg'
c*im.t^{h}oŋ.tə.wi	c*im.t^{h}oŋ	tə.wi	'sweltering heat'
ca.kal.kil	ca.kal	kil	'gravel road'
ca.kal.pat^h	ca.kal	pat^h	'a gravelly field'
ca.ri.ta.t^{h}um	ca.ri	ta.t^{h}um	'a rush for seats'
caŋ.ma.pi	caŋ.ma	pi	'monsoon'
c^{h}æk.pat^h.c^{h}im	c^{h}æk	pat^h.c^{h}im	'plastic sheet for writing'
c^{h}ap.s*al.pap	c^{h}ap.s*al	pap	'boiled glutinous rice'
c^{h}ən.tuŋ.so.ri	c^{h}ən.tuŋ	so.ri	'sound of thunder'
ci.ke.ta.ri	ci.ke	ta.ri	'legs of a frame'
ci.ne.pal	ci.ne	pal	'centipede legs'
cim.su.re	cim	su.re	'a cart'
co.kæ.tə.mi	co.kæ	tə.mi	'a pile of clams'
co.kak.pæ	co.kak	pæ	'a small boat'
com.pəl.re	com	pəl.re	'silverfish'
coŋ.i.co.kak	coŋ.i	co.kak	'a piece of paper'
coŋ.i.pæ	coŋ.i	pæ	'a paper boat'
cul.tam.pæ	cul	tam.pæ	'a chain-smoke'

cwi.paŋ.ul	cwi	paŋ.ul	‘a Dutchman’s pipe’
ə.k*æ.kit	ə.k*æ	kit	‘wing bow’
ə.k*æ.toŋ.mu	ə.k*æ	toŋ.mu	‘laying one’s hand upon a person’s shoulder’
ək.ci.so.ri	ək.ci	so.ri	‘unfair words’
əl.ɨm.cu.mə.ni	əl.ɨm	cu.mə.ni	‘an ice pack’
əŋ.təŋ.paŋ.a	əŋ.təŋ	paŋ.a	‘pratfall’
ha.nɨl.so	ha.nɨl	so	‘a longicorn beetle’
ha.nɨl.ta.ram.cwi	ha.nɨl	ta.ram.cwi	‘a flying squirrel’
him.ca.raŋ	him	ca.raŋ	‘boast of one’s strength’
him.cul	him	cul	‘tendon’
hɨrk.tə.mi	hɨrk	tə.mi	‘a mound’
hɨrk.təŋ.ə.ri	hɨrk	təŋ.ə.ri	‘a clod’
ho.pak.pəl	ho.pak	pəl	‘a carpenter bee’
hwal.si.wi	hwal	si.wi	‘a bowstring’
hyə.pa.nɨl	hyə	pa.nɨl	‘sore on a tongue’
i.ɨm.cul	i.ɨm	cul	‘slur’
i.p*al.ca.kuk	i.p*al	ca.kuk	‘teeth-marks’
i.sɨl.paŋ.ul	i.sɨl	paŋ.ul	‘a dewdrop’
i.sɨl.pi	i.sɨl	pi	‘a light rain’
il.kæ.mi	il	kæ.mi	‘a worker ant’
il.som.s*i	il	som.s*i	‘workmanship’
ip^{h}.tam.pæ	ip^{h}	tam.pæ	‘leaf tobacco’
k*a.c^{h}i.cip	k*a.c^{h}i	cip	‘magpie’s nest’
k*o.c^{h}i.ku.i	k*o.c^{h}i	ku.i	‘grilled skewers’
k*ok.tu.kak.si	k*ok.tu	kak.si	‘puppet’
k*ok.tu.sæ.pyək	k*ok.tu	sæ.pyək	‘wee hours’
k*ol.t*u.ki.cət	k*ol.t*u.ki	cət	‘salted baby octopuses’
k*ot.cul.ki	k*ot	cul.ki	‘flower stem’
k*ot.ku.kyəŋ	k*ot	ku.kyəŋ	‘flower viewing’
k*ot.pa.ku.ni	k*ot	pa.ku.ni	‘flower basket’
k*ot.pat.c^{h}im	k*ot	pat.c^{h}im	‘a calyx’
k*ot.si.kye	k*ot	si.kye	‘a floral clock’
k*ot.soŋ.i	k*ot	soŋ.i	‘blossom’

k*ot.ta.pal	k*ot	ta.pal	‘a flower bouquet’
k*ot.toŋ.san	k*ot	toŋ.san	‘a flowery hill’
k*ot.ke.cap.i	k*ot.ke	cap.i	‘fishing blue crabs’
k*ul.cu.mə.ni	k*ul	cu.mə.ni	‘a honey sac’
k*ul.tan.ci	k*ul	tan.ci	‘a jar of honey’
k*um.ca.ri	k*um	ca.ri	‘a dream’
k*um.toŋ.san	k*um	toŋ.san	‘a dreamland’
ka.cuk.sin	ka.cuk	sin	‘leather shoes’
ka.ɨl.pa.ram	ka.ɨl	pa.ram	‘an autumn wind’
ka.ræ.so.ri	ka.ræ	so.ri	‘phlegm sounds’
ka.rak.kuk.su	ka.rak	kuk.su	‘Korean vermicelli’
ka.ro.cul	ka.ro	cul	‘a horizontal line’
ka.ru.pi.nu	ka.ru	pi.nu	‘soap powder’
ka.si.təm.pul	ka.si	təm.pul	‘a bramble’
ka.sɨm.sal	ka.sɨm	sal	‘brisket’
ka.sɨm.sok	ka.sɨm	sok	‘one’s mind’
ka.sɨm.tul.re	ka.sɨm	tul.re	‘chest size’
ka.un.te.son.ka.rak	ka.un.te	son.ka.rak	‘middle finger’
kæ.ki.rɨm	kæ	ki.rɨm	‘grease on a face’
kæ.mi.cip	kæ.mi	cip	‘ant’s nest’
k^hal.ca.kuk	k^hal	ca.kuk	‘a blade scar’
k^hal.ca.ru	k^hal	ca.ru	‘hilt’
k^hal.cip	k^hal	cip	‘a cut made by a knife’
k^ho.kəl.i	k^ho	kəl.i	‘a nose pendant’
k^ho.pa.nɨl	k^ho	pa.nɨl	‘a hooked needle’
k^hoŋ.ka.ru	k^hoŋ	ka.ru	‘bean flour’
k^hoŋ.ki.rɨm	k^hoŋ	ki.rɨm	‘soybean oil’
k^hoŋ.ko.mul	k^hoŋ	ko.mul	‘powdered soybean’
k^hoŋ.kuk.su	k^hoŋ	kuk.su	‘cold bean-soup noodles’
ki.rɨm.cip	ki.rɨm	cip	‘an oil shop’
ki.rɨm.kəl.re	ki.rɨm	kəl.re	‘oilcloth’
ki.rɨm.paŋ.ul	ki.rɨm	paŋ.ul	‘oil drop’
kil.pa.tak	kil	pa.tak	‘road’
kim.ku.i	kim	ku.i	‘grilled seaweed’

kim.pap	kim	pap	‘rice rolled with seaweed’
kɨ.mul.cu.mə.ni	kɨ.mul	cu.mə.ni	‘a net pocket’
ko.c^{h}u.ki.rɨm	ko.c^{h}u	ki.rɨm	‘chili oil’
ko.ki.co.kak	ko.ki	co.kak	‘a peace of meat’
ko.ki.ku.i	ko.ki	ku.i	‘roast meat’
ko.ki.pap	ko.ki	pap	‘bait’
ko.ki.pi.nɨl	ko.ki	pi.nɨl	‘fish scales’
ko.ki.təŋ.ə.ri	ko.ki	təŋ.ə.ri	‘a meat loaf’
ko.mu.pa.kwi	ko.mu	pa.kwi	‘a rubber tire’
ko.mu.pæ	ko.mu	pæ	‘a rubber boat’
ko.ri.tan.c^{h}u	ko.ri	tan.c^{h}u	‘a buttonhook’
kol.pa.ram	kol	pa.ram	‘valley wind’
ku.ri.kə.ul	ku.ri	kə.ul	‘a copper mirror’
ku.tu.cu.kək	ku.tu	cu.kək	‘a shoehorn’
ku.tu.sol	ku.tu	sol	‘a shoebrush’
kuŋ.tuŋ.paŋ.a	kuŋ.tuŋ	paŋ.a	‘a fall on one’s buttocks’
ma.ɨm.ca.ri	ma.ɨm	ca.ri	‘nature’
ma.nɨl.caŋ.a.c*i	ma.nɨl	caŋ.a.c*i	‘pickled garilcs’
ma.taŋ.pal	ma.taŋ	pal	‘wide foot’
mal.caŋ.nan	mal	caŋ.nan	‘wordplay’
mal.pal.kup	mal	pal.kup	‘a horse’s hoof’
mal.pət	mal	pət	‘a company to talk to’
mal.so.ri	mal	so.ri	‘voice’
mal.som.s*i	mal	som.s*i	‘eloquency’
mal.ta.t^{h}um	mal	ta.t^{h}um	‘a quarrel’
mal.tæ.k*u	mal	tæ.k*u	‘a retort’
mal.toŋ.mu	mal	toŋ.mu	‘a company to talk to’
me.mil.ka.ru	me.mil	ka.ru	‘buckwheat flour’
mə.ri.ki.rɨm	mə.ri	ki.rɨm	‘hair oil’
mə.ri.sut	mə.ri	sut	‘thickness of hair’
məŋ.ca.kuk	məŋ	ca.kuk	‘bruise marks’
mit.kə.rɨm	mit	kə.rɨm	‘foundation’
mit.kɨ.rim	mit	kɨ.rim	‘a rough sketch’
mit.pa.t^{h}aŋ	mit	pa.t^{h}aŋ	‘basis’

mit.pat.c^him	mit	pat.c^him	‘an underlay’
mit.sul	mit	sul	‘crude liquor’
mok.ku.məŋ	mok	ku.məŋ	‘throat’
mok.təl.mi	mok	təl.mi	‘nape of one’s neck’
mok.tul.re	mok	tul.re	‘neck size’
mul.cam.ca.ri	mul	cam.ca.ri	‘a damsel fly’
mul.caŋ.ku	mul	caŋ.ku	‘splashing’
mul.cul.ki	mul	cul.ki	‘stream of water’
mul.kal.k^hwi	mul	kal.k^hwi	‘webbed feet’
mul.kɨ.rim.ca	mul	kɨ.rim.ca	‘shadows on the water’
mul.kɨ.rɨt	mul	kɨ.rɨt	‘a water bowl’
mul.pa.ka.ci	mul	pa.ka.ci	‘a gourd for dipping water’
mul.pi.nu	mul	pi.nu	‘liquid soap’
mul.si.kye	mul	si.kye	‘a water clock’
mul.so.ri	mul	so.ri	‘water noise’
mul.su.ce.pi	mul	su.ce.pi	‘ducks and drakes’
mul.toŋ.i	mul	toŋ.i	‘water jar’
mul.re.pa.k^hwi	mul.re	pa.k^hwi	‘the wheel of a water mill’
na.mu.cət.ka.rak	na.mu	cət.ka.rak	‘wooden chopsticks’
na.mu.co.kak	na.mu	co.kak	‘a piece of wood’
na.mu.cul.ki	na.mu	cul.ki	‘the trunk of a tree’
na.mu.ka.ci	na.mu	ka.ci	‘a twig’
nak.si.cul	nak.si	cul	‘fishing line’
nak.si.pa.nɨl	nak.si	pa.nɨl	‘a fish hook’
nak.si.pæ	nak.si	pæ	‘a fishing boat’
nat.cam	nat	cam	‘nap’
nat.ka.cuk	nat	ka.cuk	‘the skin of the face’
nat.ka.rim	nat	ka.rim	‘inhibition’
nat.sul	nat	sul	‘a drink in the daytime’
nat.to.k*æ.pi	nat	to.k*æ.pi	‘a ghost in daylight’
ne.mo.ki.tuŋ	ne.mo	ki.tuŋ	‘rectangular column’
no.ræ.ka.rak	no.ræ	ka.rak	‘melody of a song’
non.tu.rəŋ	non	tu.rəŋ	‘a levee’
non.tuk	non	tuk	‘a ridge’

nu.i.toŋ.sæŋ	nu.i	toŋ.sæŋ	‘younger sister’
nun.cu.rɨm	nun	cu.rɨm	‘eye wrinkles’
nun.ka.ru	nun	ka.ru	‘snow particles’
nun.kɨm	nun	kɨm	‘gradations on a ruler’
nun.sa.ram	nun	sa.ram	‘a snowman’
nun.sal	nun	sal	‘a frown’
nun.si.ul	nun	si.ul	‘the edge of an eyelid’
nun.soŋ.i	nun	soŋ.i	‘snowflake’
nun.tæ.cuŋ	nun	tæ.cuŋ	‘eye measurement’
nun.tu.təŋ	nun	tu.təŋ	‘an upper eyelid’
p*al.ræ.cip.ke	p*al.ræ	cip.ke	‘clothespin’
p*al.ræ.cul	p*al.ræ	cul	‘clothesline’
p*al.ræ.paŋ.maŋ.i	p*al.ræ	paŋ.maŋ.i	‘a laundry bat’
p*al.ræ.pi.nu	p*al.ræ	pi.nu	‘laundry soap’
p*u.ri.cul.ki	p*u.ri	cul.ki	‘a rhizome’
pa.ci.cə.ko.ri	pa.ci	cə.ko.ri	‘pants and jacket’
pa.k*at.sa.ram	pa.k*at	sa.ram	‘husband’
pa.nɨl.ku.məŋ	pa.nɨl	ku.məŋ	‘a hole made by a needle’
pa.nɨl.kwi	pa.nɨl	kwi	‘the eye of a needle’
pa.nɨl.paŋ.sək	pa.nɨl	paŋ.sək	‘a pincushion’
pa.ram.kæ.pi	pa.ram	kæ.pi	‘a pinwheel’
pa.ram.ku.məŋ	pa.ram	ku.məŋ	‘an air hole’
pa.ta.ka.cæ	pa.ta	ka.cæ	‘lobster’
pa.ta.pa.ram	pa.ta	pa.ram	‘sea breeze’
pa.ta.sa.ram	pa.ta	sa.ram	‘a seafarer’
pal.ca.c^{h}wi	pal	ca.c^{h}wi	‘footprints’
pal.kup	pal	kup	‘a hoof’
pam.cam	pam	cam	‘night’s sleep’
pam.son.nim	pam	son.nim	‘a night theif’
paŋ.ul.pæm	paŋ.ul	pæm	‘rattlesnake’
paŋ.ul.sæ	paŋ.ul	sæ	‘a goldfinch’
pap.cu.kək	pap	cu.kək	‘a rice paddle’
pap.cul	pap	cul	‘a source of income’
pat.c^{h}im.tol	pat.c^{h}im	tol	‘socle’

pə.sən.pal	pə.sən	pal	‘stocking feet’
p^hal.pe.kæ	p^hal	pe.kæ	‘an arm as a pillow’
p^han.so.ri	p^han	so.ri	‘Korean folk play’
p^hi.ca.kuk	p^hi	ca.kuk	‘bloodstain’
p^hi.ko.rɨm	p^hi	ko.rɨm	‘bloody pus’
p^hi.paŋ.ul	p^hi	paŋ.ul	‘beads of blood’
p^hul.cul.ki	p^hul	cul.ki	‘windlestraw’
p^hul.pəl.re	p^hul	pəl.re	‘insects’
p^hum.sak	p^hum	sak	‘wages for labor’
pi.nɨl.cul.ki	pi.nɨl	cul.ki	‘scaly bulb’
pi.pim.kuk.su	pi.pim	kuk.su	‘spicy noodles’
pi.pim.pap	pi.pim	pap	‘Bibimbap; Korean traditional dish’
pi.t^hal.pat	pi.t^hal	pat	‘hillside farm’
pit.cul.ki	pit	cul.ki	‘rain streak’
pit.tə.mi	pit	tə.mi	‘huge debt’
po.ri.ka.ru	po.ri	ka.ru	‘barley flour’
puk.so.ri	puk	so.ri	‘the sound of a drum’
pul.caŋ.nan	pul	caŋ.nan	‘playing with fire’
pul.ka.ma	pul	ka.ma	‘a hot kiln’
pul.ki.tuŋ	pul	ki.tuŋ	‘a pillar of fire’
pul.ku.təŋ.i	pul	ku.təŋ.i	‘fire pit’
put.kɨl.s*i	put	kɨl.s*i	‘calligraphy’
s*al.ka.ru	s*al	ka.ru	‘rice flour’
s*al.kyə	s*al	kyə	‘rice bran’
s*al.pap	s*al	pap	‘cooked rice’
s*ɨ.re.ki.tə.mi	s*ɨ.re.ki	tə.mi	‘a rubbish heap’
sa.i.so.ri	sa.i	so.ri	‘an epenthetic sound’
sæ.so.ri	sæ	so.ri	‘a birdcall’
sæ.k*i.cul	sæ.k*i	cul	‘a straw rope’
sæ.t^həl.ku.rɨm	sæ.t^həl	ku.rɨm	‘a cirrus’
sal.ka.cuk	sal	ka.cuk	‘the skin’
san.cim.sɨŋ	san	cim.sɨŋ	‘mountain animals’
san.pa.ram	san	pa.ram	‘a mountain wind’

san.pi.tul.ki	san	pi.tul.ki	'a turtledove'
səm.sa.ram	səm	sa.ram	'an islander'
səŋ.nyaŋ.kæ.pi	səŋ.nyaŋ	kæ.pi	'a matchstick'
si.kol.cip	si.kol	cip	'a country house'
si.kol.ku.sək	si.kol	ku.sək	'a remote rural area'
sil.pa.ram	sil	pa.ram	'a light breeze'
sil.pan.ci	sil	pan.ci	'a thin ring'
sin.pal.cu.mə.ni	sin.pal	cu.mə.ni	'a shoe pouch'
so.k*up.caŋ.nan	so.k*up	caŋ.nan	'playing house'
so.kɨm.ku.i	so.kɨm	ku.i	'salt glazed'
sol.paŋ.ul	sol	paŋ.ul	'a pine cone'
som.cə.ko.ri	som	cə.ko.ri	'a padded jacket'
som.pa.ci	som	pa.ci	'padded trousers'
som.paŋ.maŋ.i	som	paŋ.maŋ.i	'light punishment'
som.təŋ.i	som	təŋ.i	'a ball of cotton'
son.ca.kuk	son	ca.kuk	'a handprint'
son.cæ.cu	son	cæ.cu	'dexterity'
son.caŋ.nan	son	caŋ.nan	'toying with hands'
son.ka.paŋ	son	ka.paŋ	'a handbag'
son.pa.nɨ.cil	son	pa.nɨ.cil	'hand sewing'
son.pa.tak	son	pa.tak	'a palm'
son.sa.ræ	son	sa.ræ	'waving hands'
son.su.re	son	su.re	'a handcart'
sot.tan.ci	sot	tan.ci	'a caldron'
su.re.pa.k^hwi	su.re	pa.k^hwi	'a wagon wheel'
sul.caŋ.sa	sul	caŋ.sa	'a liquor business'
sul.ko.re	sul	ko.re	'a strong drinker'
sul.sim.pu.rɨm	sul	sim.pu.rɨm	'serving drinks'
sul.son.nim	sul	son.nim	'guests for a drink'
sul.toŋ.i	sul	toŋ.i	'a jar for drinks'
sun.tæ.kuk	sun.tæ	kuk	'Korean blood sausage soup'
swe.to.k*i	swe	to.k*i	'an iron ox'
t*am.cul.ki	t*am	cul.ki	'dripping perspiration'
t*am.ku.məŋ	t*am	ku.məŋ	'sweat pore'

t*am.paŋ.ul	t*am	paŋ.ul	‘a drop of perspiration’
t*am.pəm.pək	t*am	pəm.pək	‘sweating like a pig’
t*aŋ.kə.mi	t*aŋ	kə.mi	‘dusk’
t*e.to.tuk	t*e	to.tuk	‘a gang of robbers’
t*oŋ.pæ.c*aŋ	t*oŋ	pæ.c*aŋ	‘reckless courage’
tak.coŋ.i	tak	coŋ.i	‘Dakpaper; Korean traditional paper’
tal.təŋ.i	tal	təŋ.i	‘a full moon’
tal.toŋ.ne	tal	toŋ.ne	‘poor hillside village’
talk.sal	talk	sal	‘goose bumps’
talk.ta.ri	talk	ta.ri	‘a drumstick’
talk.tæ.ka.ri	talk	tæ.ka.ri	‘bird-brain’
tam.pyə.rak	tam	pyə.rak	‘wall’
tan.kol.cip	tan.kol	cip	‘a favorite hangout’
te.ril.sa.wi	te.ril	sa.wi	‘a son-in-law who lives with his wife’s family’
t^həl.ka.cuk	t^həl	ka.cuk	‘pelage’
t^həl.ku.məŋ	t^həl	ku.məŋ	‘a pore of the skin’
tɨl.ki.rɨm	tɨl	ki.rɨm	‘perilla seed oil’
to.k*æ.pi.pul	to.k*æ.pi	pul	‘spunkie’
tol.cəl.ku	tol	cəl.ku	‘a stone mortar’
tol.co.kak	tol	co.kak	‘a stone fragment’
tol.cu.mə.ni	tol	cu.mə.ni	‘a stone pouch’
tol.ka.ru	tol	ka.ru	‘stone dust’
tol.ki.tuŋ	tol	ki.tuŋ	‘a stone pillar’
tol.ki.wa	tol	ki.wa	‘a slate’
tol.pe.kæ	tol	pe.kæ	‘a stone as a pillow’
tol.sot	tol	sot	‘a stone pot’
tol.ta.ri	tol	ta.ri	‘a stone bridge’
tol.tæ.ka.ri	tol	tæ.ka.ri	‘bonehead’
tol.tam	tol	tam	‘a stone wall’
tol.to.k*i	tol	to.k*i	‘a stone ax’
ton.caŋ.sa	ton	caŋ.sa	‘a moneylending business’
ton.paŋ.sək	ton	paŋ.sək	‘a pile of money’
ton.pəl.re	ton	pəl.re	‘a millipede’

ton.pyə.rak	ton	pyə.rak	‘a jackpot’
twæ.ci.ki.rɨm	twæ.ci	ki.rɨm	‘lard’
u.sɨm.pa.ta	u.sɨm	pa.ta	‘howls of laughters’
u.sɨm.po.t*a.ri	u.sɨm	po.t*a.ri	‘a bundle of laughs’
u.sɨm.so.ri	u.sɨm	so.ri	‘sound of laughter’
ul.rim.so.ri	ul.rim	so.ri	‘ringing sound’
waŋ.pa.ram	waŋ	pa.ram	‘a violent storm’

References

Ahn, Pyeong-Hee (1968). On the genitive suffix '-s' in Middle Korean [Cuŋse kukəui sokkyək əmi '-s'e tæhayə]. *In Festschrift for Professor Sungnyeong Lee*. Seoul: Ulyu Munhwasa. 337–345.

Ahn, Sang-Cheol (1985). *The interplay of phonology and morphology in Korean*. PhD dissertation, University of Illinois, Urbana-Champaign.

Aum, Tae-Su (1998). The study of Korean tensification phenomena. *Korean Language and Literature in International Context* **19**. International Association of Language and Literature. 3–17.

Bates, Douglas M., Martin Maechler, Ben Bolker & Steven Walker (2014). Package 'lme4': linear mixed-effects models using Eigen and S4. Version 1.0-6. http:// cran.r-project.org/web/packages/lme4/index.html.

Byeon, Yong-Woo (2012). A Study on the fortition in the back of the voiced consonants: related to set the rule of 't' insertion. *The Research on Korean Language and Literature* **59**. The Association of the Research on Korean Language and Literature. 223–263.

Choi, Hyon-Pai (1955). *The grammar of our language* [Urimalpon]. Seoul: Cəŋɨmsa.

Chung, Kook (1980). *Neutralization in Korean: a functional view*. PhD dissertation, University of Texas at Austin. Seoul: Hanshin.

Coetzee, Andries (2004). *What it means to be a loser: Non-optimal candidates in Optimality Theory*. PhD dissertation, University of Massachusetts, Amherst.

Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Andres Eriksson & Östen Dahl (eds.) *Proceedings of the Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111–120.

Ha, Se-Kyeong (2006). *The morphology and phonology of 'sai-sori' phenomena in modern Korean*. PhD dissertation, Seoul National University.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* **39**. 379–440.

Hayes, Bruce (2009a). Maxent Grammar Tool. Software package. http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/

Hayes, Bruce (2009b). *Introductory Phonology*. Wiley-Blackwell. Chichester.

Huh, Woong (1986). *Korean linguistics* [Kukəhak]. Seoul: Sæmmunhwasa.

Im, Hong-Pin (1981). To solve the problems on *saisiot* [Saisiot munceui hækyəlɨl wihayə]. *Kukehak* **10**. The Society of Korean Linguistics. 1–35.

Ito, Chiyuki (2014). Compound tensification and laryngeal co-occurrence restrictions in Yanbian Korean. *Phonology* **31**(3). 349-398.

Ito, Junko & Armin Mester (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Language Inquiry* **17**. 47–73.

Ito, Junko & Armin Mester (1998). Markedness and word structure: OCP effects in Japanese. Ms. University of California, Santa Cruz.

Jun, Sun-Ah (1993). *The phonetics and phonology of Korean prosody.* PhD dissertation, Ohio State University.

Jun, Sun-Ah (1998). The accentual phrases in the Korean prosodic hierarchy. *Phonology* **15**(2). 189–226.

Kang, Beom-Mo & Hung-Gyu Kim (2009). *Korean usage frequency* [Hankukə sayoŋ pinto]. Institute of Korean Culture.

Kawahara, Shigeto & Shin-ichiro Sano (2014). Testing Rosen's Rule and Strong Lyman's Law. *NINJAL Research Papers* **7**. 111–120.

Kim, Cha-Kyun (1992). Phonology of Epenthetic /S^h/. *Journal of Korea Linguistics* **22**. The Society of Korean Linguistics. 191–236.

Kim, Chang-Sop (1994). *Word formation and word structure in Korean*. PhD dissertation, Seoul National University.

Kim, Chin-Wu (1970). Boundary phenomena in Korean. *Paper in Linguistics* **2**. 1–26.

Kim, Choong-Bae (1974). Tensification revisited. *Language Research* **10**(2). Language Education Institute. 129–142.

Kim-Renaud, Young-Key (1974). *Korean consonantal phonology*. PhD dissertation, University of Hawaii.

Kim, Seon-Cheol (2003). *A Study of the Current Use of Standard Pronunciations* [Pyocun Paləm Siltæ Cosa]. The National Institute of the Korean Language [Kuklip kukə yənkuwən].

Kirchner, Robert Martin. (1998). *An effort-based approach to consonant lenition*. PhD dissertation, UCLA.

Ko, Kwang-Mo (1996). Two Sound Changes Related to r in Korean. *Eoneohag* **18**. The Linguistic Society of Korea. 31–50.

Kuklip kukə yənkuwən (1999). *Standard Korean dictionary* [Pyocun kukə tæsacən]. Seoul: Tusantoŋa.

Kurisu, Kazutaka (2001). *The phonology of morpheme realization*. PhD dissertation, University of California, Santa Cruz.

Labrune, Laurence (1999). Variation intra et inter-langue: morpho-phonologie du rendaku en japonais et du sai-sios en coréen. *Cahiers de Grammaire* **24**. 117–152.

Lee, Chung-Min (1972). Boundary phenomena in Korean revisited. *Papers in Linguistics* **5**. 454–474.

Lee, Ho-Young (1996). *Korean Phonetics* [Kukəɨmsəŋhak]. Seoul: Thaehaksa.

Lee, Ho-Young (2009). Tensification preference of native Seoul speakers of Korean. *Journal of The Korean Society of Speech Sciences* **1**(2). The Korean Association of Speech Sciences. 151–162.

Lyman, Benjamin S. (1894). Change from surd to sonant in Japanese compounds. *Oriental Studies of the Oriental Club of Philadelphia*. 160–176.

Martin, Andrew (2011). Grammars leak: modeling how phonotactic generalizations interact within the grammar. *Language* **87**(4). 751–770.

McCarthy, John J. (1986). OCP effects: gemination and antigemination. *Linguistic Inquiry* **17**. 207–263.

Moon, Yang-Soo (1997). The ‘Sais-sori’ Phenomena of Korean. *Journal of Humanities* **37**. 63–80.

Nusbaum, Howard C., David B. Pisoni & Christopher K. Davis (1984). Sizing up the Hoosier mental lexicon: measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report* **10**. Indiana University. 357–375.

Oh, Jung-Ran (1987). Fortition within Korean compounds. *Korean Journal of Linguistics* **12**(1). 35–53.

Oh, Jung-Ran (1988). *A historical study on Korean tense consonants* [kyəŋɨmui kukəsacək yənku]. Seoul: Hanshin.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.

R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Ramstedt, Gustav J. (1939). *A Korean Grammar*. Helsinki: Société Finno-Ougrienne.

Ryu, Chang-Don (1963). On the additional /t/ phenomenon. *The Journal of Korean Studies* **7**. Yonsei University. 1–19.

Shim, Jae-Kee (1979). Semantic functions of modifications in Korean. *Language Research* **15**(2). Seoul National University Language Education Research Center. 109–121.

Shin, Ji-Young (1999). Prosodic units and tensification in Korean. *Korean Linguistics* **10**. The Association for Korean Linguistics. 5–25.

Vance, Timothy (1987). *An Introduction to Japanese Phonology*. Albany: State University of New York Press.

Yi, Bong-Won (2002). *A usage-based approach to Korean phonology*. PhD dissertation, Korea University.

Zipf, G. K. (1935). *The psycho-biology of language: an introduction to dynamic philology*. Boston: Houghton Mifflin company.

Zuraw, Kie (2000). *Patterned exceptions in phonology*. PhD dissertation, UCLA.

Zuraw, Kie (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory* **28**(2). 417–472.

Zuraw, Kie (2011). Predicting Korean sai-siot: phonological and non-phonological factors. Handout from paper presented at the 21st Japanese/Korean Linguistics Conference. Seoul National University.

국문초록

서울 방언 합성어 경음화의 음운론적 경향

본 논문의 목적은 서울 방언의 합성어에서 수의적으로 발생하는 경음화 현상의 음운론적 경향을 밝히고, 최대 엔트로피 최적성 이론 문법을 기반으로 한 형식적 분석을 제시하는 것이다.

이를 위해 우선 서울 방언 화자를 대상으로 설문조사를 실시하여 대량의 발음 자료를 얻어, 환경에 따라 경음화의 정도가 어떻게 달라지는지 확인하였다. 설문조사 결과에서 확인된 경음화의 경향은 다음과 같다. 먼저, 경음화는 해당 합성명사의 빈도가 높을 경우 더 잘 일어난다. 또한, 첫 번째 명사의 마지막 음절 종성이 장애음일 때 가장 잘 일어나고, 비음, 유음의 순서로 경음화의 정도가 점차 낮아지며, 첫 번째 명사의 마지막 음절이 모음으로 끝날 경우에 경음화가 가장 덜 일어난다. 또한, 만약 두 번째 명사의 첫 음절 두음이 설정음이고, 동시에 첫 번째 명사의 마지막 음절 종성이 유음이면 경음화가 더 잘 일어난다. 마지막으로, 두 번째 명사가 이미 경음이나 격음을 포함하고 있다면 경음화가 저지된다.

본 연구에서는 이러한 경향성을 최적성 이론의 틀 안에서 형식화하는 것 또한 시도하였다. 우선 경음화의 발생에 대해서는, 경음화를 야기하는 제약인 REALIZEMORPHEME과 경음화를 방지하는 제약인 IDENT(tense)를 각각 설정하였다. 설문조사 결과에서 나타난 여러가지 경향성을 설명하기 위한 제약들도 각각 설정하였다. 먼저 합성명사의 빈도에 관해서는, 낮은 빈도의 경음화를 막는 제약인 *TENSE/LOWFREQUENCY 를 설정하였다. 첫 번째 명사의 마지막 음절 종성이 장애음일 때 경음화의 정도가 가장 높았던 것은, 한국어의 강세구 내에서 필수적으로 적용되는 음운규칙인 장애음 뒤 경음화가 합성어의 단어 경계에도 작용한 결과라고 보았다. 따라서 강세구 내에서 장애음 뒤에 평음이 오는 것을 방지하는 *obs-lax 제약을 설정하였다. 한편 Ito (2014)는 선행 분절음에 따라 경음화가 달라지는 것이 연음화와 관련이 있다고 논의하였다. 즉, 비음, 유음과 모음 뒤에서 각각 다른 정도로 경음화가 나타나는 것은 각 환경에 따라 연음화를 선호하는 정도, 혹은 경음화를 방지하는 정도가 달라지기 때문이라는 것이다. 이를 바탕으로 각 환경에서 경음화를 방지하는 제약 *TENSE/VOWEL_, *TENSE/LIQUID_와 *TENSE/NASAL_를 설정하였다. 또한, 첫 번째 명사의 마지막 음절 종성이 유음이고, 동시에 두 번째 명사의 첫 음절 초성이 설정음일 경우 경음화가 더 많이 일어나는 경향은 한국어의 음소배열제약이 영향을 끼친 것으로 보았다. 한국어 고유어의 형태소 내부에서 유음 뒤에 오는 설정음은 반드시 경음이다. 즉, Martin (2011)에서 증명되었던 바와 같이, 형태소 내부에서만 작용하는 것으로 여겨졌던 음소배열제약이 형태소의 경계를 넘어 합성어의 단어 경계에까지 그 영향을 미친 결과라는 것이다. 이를 포착하기 위해 유음

뒤에서 평음인 설정음의 출현을 방지하는 제약인 *L(+)C 를 설정하였다. 마지막으로, 두 번째 명사가 이미 경음이나 격음을 포함하고 있을 때에 경음화가 저지되는 경향은, Ito (2014)에서 밝힌 바와 같이 후두 자질의 공기 제약이 그 원인이라고 보았다. 따라서 한 단어 내부에서 후두 자질의 동시출현을 막는 제약인 OCP(stem)을 설정하였다.

본 연구에서는 서울 방언 합성어 경음화 현상의 수의성을 포착하기 위해 확률적 최적성 이론 문법의 하나인 최대 엔트로피 모형을 채택하였다. 그리고 설문조사 결과에서 확보한 환경별 경음화 정도 자료를 바탕으로 위 제약들의 구체적인 가중치를 구하였다. 또한 계산된 제약 가중치를 통해 가상학습을 실시한 결과, 최대 엔트로피 모형에 의해 설정한 문법이 실제 자료에서 나타난 경음화의 환경별 분포를 성공적으로 포착함을 확인하였다.

주요어: 합성어 경음화, 사이시옷, 비범주적 음운 현상, 음운론적 경향, 확률적 최적성 이론, 최대 엔트로피 모형

학 번: 2014-20106