Unnatural phonology may occur as the result of grammatical leakage

Unnatural phonological patterns exist, although they are rare across languages. Typically, these patterns cannot be described in terms of phonetically motivated processes such as assimilation. The most widely accepted explanation for such unnatural patterns is that they result from a series of historical changes, each of which is natural, but which collectively obscure phonetic motivation (Blevins 2004, 2008; Hayes et al. 2009; and others). In this study, I argue that at least some unnatural patterns are the result of a Maximum Entropy (MaxEnt) learning algorithm with a Gaussian prior. Specifically, I present a MaxEnt modeling of the historical development of Korean preliquid nasalization, demonstrating that it is a case of grammatical leakage, where a generalization that holds for a specific grammatical domain or environment leaks into a more general one (Martin 2011).

In Seoul Korean, an obstruent becomes nasalized before a nasal, as illustrated in (1). This obligatory obstruent nasalization triggered by a following nasal is a common type of assimilation observed cross-linguistically. Seoul Korean also exhibits an additional type of obstruent nasalization, triggered by a following liquid, as seen in (2). The traditional literature (e.g., Kim-Renaud 1974) argues that a post-consonantal liquid is obligatorily nasalized, which then feeds prenasal nasalization. Under this account, Korean preliquid nasalization would simply be the combined effect of two independent natural constraints, *ONSETL ("No liquids in the onset") and *OBSNAS ("No obstruents before a nasal"). However, as is experimentally shown by Jun (2000), the nasalization of a post-consonantal liquid is optional in Seoul Korean, and the preceding stop becomes nasalized regardless of whether the following liquid is nasalized or not, as shown in (2). Since a following *non-nasal* liquid is clearly a phonologically unnatural environment for obstruent nasalization, and preliquid nasalization, as observed in Contemporary Seoul Korean, is at best rare, if not almost completely absent, across languages, it can be considered a case of unnatural phonology.

I analyze preliquid nasalization as a result of the MaxEnt learning process for prenasal nasalization. In this analysis, I hypothesize the historical stages of Seoul Korean, shown in (3), based on previous research on Korean phonology and morphology. In the earlier stages 1-2, obstruent nasalization occurs exclusively before a (surface) nasal, and a constraint prohibiting obstruent codas before nasal onsets (*OBSNAS) adequately accounts for the observed patterns. However, in stage 3, which corresponds to contemporary Seoul Korean, a large number of recent loanwords, primarily from English, with initial liquids are borrowed into Korean. Consequently, a constraint prohibiting liquids in the onset (*ONSETL) weakens, allowing liquids in the onset to optionally surface as such. Obstruent nasalization is now triggered not only by surface nasals but also by surface liquids, indicating the activation of a general constraint prohibiting obstruent codas before sonorant consonant onsets (*OBSSON). I argue that the emergence of *OBSSON in Contemporary Seoul Korean is an instance of grammar leakage, where a generalization that holds for a specific phonological environment, i.e., before a nasal, leaks into a more general one, i.e., before a sonorant.

Assuming that Korean speakers acquire a constraint-based grammar using a MaxEnt learning algorithm with a Gaussian prior, I conducted a simulation of multigenerational learning by manipulating the relative frequency of output candidates for words with a word-initial liquid (/la/) in the learning data. Specifically, I gradually increased the frequency of its faithful output candidate [la], while simultaneously decreasing the frequency of its altered competitor [na]. This simulation crudely models the gradual influx of liquid-initial loanwords into Korean. I adopted the constraints shown in (4a), with prior mu and sigma^2 values of 0 and 500, respectively. The learning data

consists of words with prenasal nasalization (/ap-na/ → [amna]) and preliquid nasalization (/ap-la/ → [amna]), in addition to those with an underlying initial liquid (/la/ → [na] ~ [la]). What is most important is that no words with nasalization before a surface liquid (/ap-la/ → [amla]) are included in the learning data.

The results indicate that as the percentage of faithful realization of a word-initial liquid increases, the weight of *OBSSON also increases, as shown in (4a). Consequently, the probability of preliquid nasalization (/ap-la/ → [amla]) also increases, as observed in (4b). Thus, this simulation illustrates how preliquid nasalization may emerge and become more frequent over time. It is noteworthy that *OBSSON can be learned with a relatively higher weight even in the absence of learning data containing words with obstruent nasalization before a surface liquid. This effect is due to the Gaussian prior, which evenly distributes the weights of the relevant constraints, whether specific (*OBSNAS) or general (*OBSSON).

In summary, I have demonstrated that unnatural phonology can emerge during the process of learning a MaxEnt grammar through computational modeling of the historical development of Korean preliquid nasalization.

(1) Prenasal nasalization in Seoul Korean

a.	/mək-ne/	[məŋne]	‘I see he is eating’
b.	/ip-mɛ/	[immɛ]	‘shape of the mouth’

(2) Preliquid nasalization in Seoul Korean

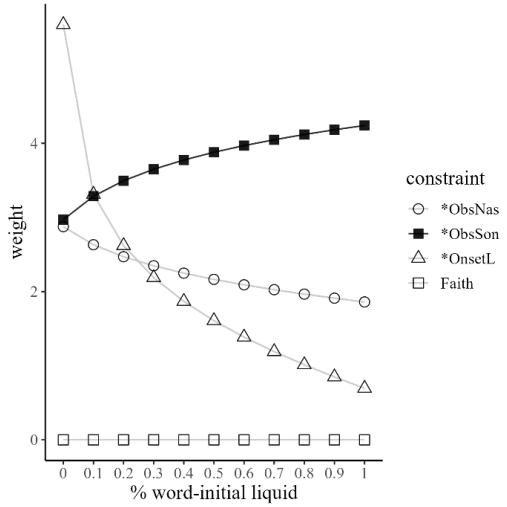
a.	/tɕhak-lo/	[tɕhaŋno] ~ [tɕhaŋlo]	‘university street’
b.	/kuk-lip/	[kuŋnip] ~ [kuŋlip]	‘national’

(3) Historical stages of Seoul Korean (n/a = not available, C = Consonantal)

stages	lexicon	/l/-initial morpheme	post-C /l/ nasalization	nasalization before a surface liquid	nasalization before a nasal
1	native Korean	no	n/a	n/a	obligatory
2	*plus* Sino-Korean	yes	obligatory	n/a	obligatory
3	*plus* recent loanwords	yes	optional	obligatory	obligatory

(4) Results of the simulation

a. constraint weight



b. probability