Reanalysis and Regularization in Morphophonology: Korean Noun Inflection and Beyond

Brent de Chene, Waseda University (Emeritus)

The University of Hong Kong, October 25, 2018

1. Introduction: the phonological life cycle and dedicated word-level phonological rules

The life cycle of phonological rules or processes is a concept that has been familiar to phonologists since the work of Kruszewski and Baudouin de Courtenay in the last decades of the 19th century. Coming into being with the phonologization (Hyman 1976) of mechanically determined phonetic variation and being as a result entirely general at the outset, phonological processes subsequently acquire morphosyntactic conditioning or come to apply in increasingly restricted morphological domains (Bermúdez-Otero 2007:504, 2015:382-384). Eventually, they may disappear from the grammar altogether, leaving traces of their former existence only in isolated alternations between lexically listed items; this is the case, for example, for the residue of Verner's Law in English (*was/were, lose/forlorn*).

At the same time, it has traditionally been recognized (see e.g. Kiparsky 1972:280) that phonological rules may originate not in the phonologization of phonetic variation, but in the reanalysis of existing phonological patterns. In particular, I claim, this is true of a class of word-level phonological rules of inflection. I will call such rules, which do not result from domain narrowing (Bermúdez-Otero 2015:382) operating on phrase-level rules, "dedicated word-level phonological rules". In addition to resulting from reanalysis, such rules, in establishing innovative generalizations, tend to render irregular a wide range of inflected forms, necessitating a long period of regularization in order to eliminate items which do not conform to them. Their diachronic profile can thus be characterized as consisting of reanalysis followed by regularization.

Dedicated word-level phonological rules have received little attention in the recent phonological literature. This is probably because they fit badly with current theoretical assumptions. In particular, in addition to failing to exemplify the standard phonological life cycle, they pose certain challenges for Optimality Theory, the dominant phonological framework since the early 1990s. To begin with, of course, they are prima facie examples of language-specific rules, a class of objects that orthodox OT claims not to exist. In addition, OT appears to provide no natural place for the principles of reanalysis necessary to explain the origin of such rules, and, as we will see below, it provides no natural treatment of regularization, either.

Dedicated word-level rules, however, are of inherent interest because of their capacity to fundamentally reshape inflectional systems. For example, in a number of Japanese and Okinawan dialects, an innovative *r*-epenthesis rule applying intervocalically at verb stem boundary has resulted, after the reinterpretation of epenthetic *r* as belonging to stems, in the total elimination of historical vowel-stem inflection in favor of the inflectional pattern of *r*-stems (de Chene 2016, 2018). In addition, dedicated word-level rules provide a window on the concept "natural phonological rule" that is independent of sound change and its roots in essentially mechanical considerations of articulation and perception.

In this talk, after preliminary comments on the nature of regularization, I will explore a case of reanalysis and regularization involving a dedicated word-level rule that has been approached from a number of angles in the recent literature, the case of Korean noun inflection. After introducing an analysis couched in terms of reanalysis and regularization whose essential features go back to Ko (1989), I will take up alternate interpretations of the variation and change in question and argue that they are unsatisfactory by comparison with the reanalysis and regularization account. In particular, I will consider accounts that attribute the variation in question to variable assignment of underlying phonological forms, to stochastic or probabilistic rules, and to reranking of Optimality Theoretic constraints. I will conclude by briefly noting several other examples of dedicated word-level rules and asking what we can learn from them both about speakers' choice of underlying representations and about phonological naturalness, noting convergence with the results of phonological experimentation on the latter point.

2. Background: the nature of regularization

"Regularization" is a concept that is commonly appealed to when, over time, the inflection of one or more items, hitherto arbitrary or unpredictable in some respect, assimilates to an unmarked or default pattern. Thus, when the conjugational pattern *help holp holpen* is replaced by *help helped, helped,* a change that was completed in the 18th century, we say that the inflection of *help* has been regularized or has undergone regularization. How, exactly, should we characterize regularization and the processes that bring it about? There are two senses, distinct but closely related, in which that question can be answered, the psychological and the grammatical.

In producing inflected forms, speakers may either retrieve them from memory or generate them using the rules of the grammar (see e.g. Albright 2002:10). Irregular forms, by definition, cannot be produced by the grammar, so they must be retrieved from memory. If retrieval of an irregular form fails because it is "not learned, remembered, or accessed fast enough" (Garrett 2008:128), speakers will either revert to a default form or derive a substitute form with the rules of the grammar. In psychological terms, then, regularization can be characterized as the result of failure to retrieve irregular forms from memory.

By the same token, since irregular forms cannot be produced by the grammar, they must be coded in the lexicon in some form or other. The psychological characterization of regularization just suggested appears to entail that irregular allomorphs or inflected forms must be lexically listed in full, but there are other possibilities for lexical coding as well. In particular, an exception feature may code the fact that some lexical item is irregular in the sense of failing to undergo a rule of the grammar. Similarly, if we admit the concept "minor rule", a rule that applies only to items specifically marked to undergo it, a minor rule feature may code irregularity as well. In grammatical terms, then, regularization will take the form of simplification of lexical entries by loss of the lexical coding of irregularity, whether that coding takes the form of a listed allomorph, an exception feature, or a minor rule feature.

A quick example: in Proto-Germanic (see Ringe 2006:284), adjectival Comparatives and Superlatives were formed either with front-vowel suffixes *-iz-, -is-tə-* (of clear IE provenance) or back-vowel suffixes *-ōz-, -ōs-tə-* (a specifically Germanic development); the former, but not the latter, conditioned stem umlaut. In pre-OE, a complex series of changes resulted in loss of suffixal vowels in the Comparative (see Ringe and Taylor 2014:273-274, 388); in the Superlative, while back-vowel suffixes were generalized in OE, the suffixal vowel was then subject to reduction (Campbell 1959:158, 272-273). As a result of these changes, the umlaut alternation was rendered opaque.

Leveling of the umlaut alternation in Comparatives and Superlatives had begun already in pre-OE (Ringe and Taylor 2014:388), and continued into the historical period. According to the OED, the alternation lasted until about 1500 in the paradigm of *strong* (Comparative *strenger*) and until shortly before 1600 in the paradigm of *long* (Comparative *lenger*). In the case of *old,* the alternation is still current in residual form, since conservative *elder/eldest* is possible as a stylistically marked option in attributive use when modifying a [+human] noun (*the elder of the two brothers/*manuscripts; *my brother is elder than I am*).

With respect to the umlaut alternation in adjectival inflection, English speakers have clearly taken failure to alternate as the regular or default pattern. Correspondingly, the lexical representation of alternating adjectives will have included either a feature triggering a minor rule of umlaut or an irregular umlauted allomorph, as in the sample entry for alternating *long* in (1) (where the specification [+Comparative] is shared by the Comparative and the Superlative).

(1) $\left[\begin{array}{l} \text{lɛŋg / \_\_ [+Comparative]} \\ \text{lɑŋg} \end{array}\right]$

Leveling will then have been the result of loss of the minor rule feature or the marked allomorph from lexical entries.

In contrast to English speakers, German speakers have, for monosyllabic back-vowel adjectives, taken the alternating

pattern as regular, and the umlaut alternation has been extended to all such adjectives even when, like *arm* 'poor', they originally took back vowel suffixes (see Ringe 2006:284) in the Comparative and Superlative. In German, then, nonalternating adjectives will have been marked with an exception feature for a (major) rule of umlaut, and extension of the alternation will have resulted from loss of this feature from lexical entries. The umlaut example illustrates, then, that both leveling and extension of an alternation are naturally conceptualized as regularization—that is, as simplification of lexical entries by loss of the lexical coding of irregularity.

In closing this section, let us note the typical diachronic profile of regularization. Regularization normally involves a long period of variation or competition between conservative and innovative forms, and this variation displays diachronic directionality or telicity, with the ratio of conservative to innovative forms decreasing, evidently monotonically, over time. For example, recorded variation between *older* and *elder* appears to date back at least as far as the year 1200, with attestation of *aldere* in place of the expected *eldere* (OED). Four hundred years later, in the works of Shakespeare, the ratio of conservative to innovative forms remains high, roughly 9 to 1, and four hundred years after that, the conservative forms, while highly restricted in distribution, have still not disappeared entirely.

The typical diachronic profile of regularization can be explained if we assume that there is a tendency for speakers to simplify lexical entries by failing to retrieve "excess information", where the latter term refers to any information over and above a simple pairing of sound and meaning, in particular including exception features, minor rule features, and listed irregular allomorphs. If, for a given lexical item, each generation of speakers increases slightly the degree to which excess information fails to be retrieved from that item's lexical entry, the ratio of conservative to innovative forms involving that item will decrease monotonically over time. This diachronic profile should be considered an idealization; as is well known, extremely common irregular items may resist regularization indefinitely, and the course of regularization may also be slowed or even reversed by normative pressures.

3. Reanalysis and regularization in Korean noun inflection

3.1 Background: Coda neutralization and cluster reduction

Where [^{h}] represents aspiration (generally analyzed as [+Spread Glottis]), ['] represents "tenseness" (generally analyzed either as [+Constricted Glottis] or as phonological gemination), and [c] is an alveopalatal affricate, Korean has the following 14 obstruents: /p p^h p' t t^h t' c c^h c' k k^h k' s s'/. In syllable codas, however, an exceptionless constraint limits obstruents to the set /p t k/. Call /p t k/ "unmarked obstruents" and all others "marked obstruents".

While only unmarked obstruents occur syllable-finally, a number of verb stems and noun stems apparently end in marked obstruents. Consider what happens in the conjugation of a verb stem ending in such an obstruent, taking as an example /coc^h-/ 'follow'. Before an ending beginning with a consonant, stem-final /c^h/ will be syllable-final, and it will reduce to [t] in order to satisfy the restriction limiting coda obstruents to /p t k/; the process that changes /c^h/ to [t] can be called "Coda Neutralization" (CN). Taking account of another automatic process that tensifies an obstruent immediately following a stop, /coc^h-ko/, a conjunctive form of the verb sometimes called the "Gerund", surfaces as [cot.k'o], where the period represents syllable boundary.

Before an ending beginning with a vowel, however, a stem-final marked obstruent like the /c^h/ of /coc^h-/ is resyllabified into the onset of the following syllable, thereby bleeding CN. For example, /coc^h-a/, another conjunctive form of the verb sometimes called the "Infinitive", surfaces as [co.c^{h}a] (the Infinitive suffix is *-a* after *a* or *o* in the previous syllable and *-ə* otherwise). A stem like /coc^h-/ thus alternates as a result of CN, ending in [c^h] before a vowel but in [t] before a consonant. Verb stems ending in other marked obstruents show parallel alternations. Alternations due to CN are stable in verb inflection, showing no variation and no tendency to change.

Korean verb and noun stems may also end in consonant clusters, and these undergo reduction to single consonants when they occur syllable-finally. In Seoul Korean, Cluster Reduction (CR) deletes the second of two consonants except when that is a labial or velar stop (in practice, one of /p p^h k m/). The verb stem /ɨlp^h-/ 'recite' thus has the

Gerund [ɨp.k'o], in whose derivation CR feeds CN, and the Infinitive [ɨl.p^{h}ə], with resyllabification of the stem-final cluster's second member. Sohn (1999:172), however, notes that younger speakers may fail to apply CR in clusters consisting of *l* and an oral stop, citing the form [ɨlp.k'o] for the Gerund of 'recite'. As that example illustrates, there is dialectal and idiolectal variation associated with CR.

3.2 Variation and change in noun inflection

One might expect that noun stems ending in marked obstruents and in clusters would alternate in the same way as verb stems depending on whether the following clitic, typically a case particle or the copula, begins with a consonant or with a vowel. This expectation is satisfied before consonant-initial clitics: CN guarantees that in that context, an obstruent at the end of a noun stem must be one of the set /p t k/. For example, the noun stems 'knee' and 'kitchen', which historically have the forms **murɨp^h* and **puək^h*, respectively, show the expected alternants [murɨp] and [puək] before a consonant-initial clitic: the results of combining these stems with /to/ 'also' surface as [mu.rɨp.t'o] and [pu.ək.t'o]. Since noun stems, unlike verb stems, can appear in isolation—that is, with zero inflection—they also show that alternant word-finally.

Unexpectedly, however, the neutralized alternants [murɨp] and [puək] may also occur before a vowel-initial clitic. Thus the results of combining the stems in question with the accusative clitic *-il* may be either the conservative forms [mu.rɨ.p^{h}ɨl] and [pu.ə.k^{h}ɨl], as the verbal examples discussed above would lead us to expect, or innovative [mu.rɨ.bɨl] and [pu.ə.gɨl], where the final segments of [murɨp] and [puək] have undergone a further automatic process voicing plain stops when both the preceding and the following sounds are voiced. In the same way, the accusative of **pak*' 'outside' varies between conservative [pa.k'ɨl] and innovative [pa.gɨl], and stems ending historically in clusters show parallel variation: [kap.sɨl] ~ [ka.bɨl] 'price (acc.)' (**kaps*), [nək.sɨl] ~ [nə.gɨl] 'soul(acc.)' (**nəks*), [hɨl.kɨl] ~ [hɨ.gɨl] 'soil (acc.)' (**hɨlk*).

What accounts for the occurrence of neutralized forms like [murɨp] and [puək] before a vowel-initial clitic? One possibility (R. Bermúdez- Otero, p.c.), in line with the fact that phonological processes typically undergo "domain narrowing" in the course of their history, is that CN is gradually coming to be applicable in the phonological domain defined by a noun stem, so that it applies in that domain regardless of what follows. That this is unlikely to be the correct explanation, however, is shown by the behavior of stems ending in coronal obstruents.

For example, the name of a popular traditional game played with four sticks (below, "yut") is historically **juc^h*, so that on the domain narrowing account of variation in prevocalic stem allomorphs we would expect the corresponding accusative to vary between [ju.c^{h}ɨl] and [ju.dɨl], with the [d] of the latter reflecting automatic voicing of the [t] that results from CN applying in the domain of the noun stem. In fact, however, accusative [ju.c^{h}ɨl] varies not with [ju.dɨl], but with [ju.sɨl]. In the same way, the accusative [pa.t^{h}ɨl] of **pat^h* 'field' varies with [pa.sɨl] rather than with [pa.dɨl], and the accusative [na.ʤɨl] of **nac* 'day' varies with [na.sɨl] rather than with [na.dɨl].

In sum, while there are also minority tendencies that we will return to below in section 4.2, the alternations *c^h/t, t^h/t, and c/t* at the end of noun stems, where in each case the first alternant appears prevocalically and the second elsewhere, are all in variation with and tending to be replaced by the alternation *s/t.* For stems in **t* that would originally have shown only the automatic voicing alternation *d/t,* this process of replacement is already complete (Kang 2003:129-130, Ito 2010:363), with the one historical **t*-stem claimed by Ito for Modern Korean, *nat* 'grain', no longer being a free noun (Martin 1992:108). It is clear, then, that the hypothesis that CN is undergoing domain narrowing, which appeared to account for the occurrence of the neutralized alternants [murɨp] 'knee' and [puək] 'kitchen' in prevocalic position, makes incorrect predictions about the development of coronal-final stems.

3.3 A reanalysis-plus-regularization account

Given the concept of regularization that we developed in section 2, all of the data we have just seen in section 3.2, as

well as a good deal of further data that will be introduced below, are accounted for by the analysis in (2).

(2) a. Default forms of noun stems coincide with their pre-consonantal and word-final allomorphs.
    b. There is a rule taking *t* to *s* when it is both syllable-initial and final in a noun stem: t → s / . __$_{N}$]

To see this, first note that as a corollary of (2a), prevocalic allomorphs of noun stems, in contrast to default forms, will be listed with their environments. The lexical representations of 'knee', 'kitchen', and 'yut', for example, will be as in (3), parallel to (1) above (these representations could be simplified by the use of angled brackets).

(3) a. { murɨp^{h} / __V ; murɨp }   b. { puək^{h} / __V ; puək }   c. { juc^{h} / __V ; jut }

Given the tendency, assumed above, for speakers to simplify lexical entries by failing to retrieve information in excess of a simple pairing of sound and meaning, the prevocalic allomorphs of (3) will be subject to gradual loss. For (3a) and (3b), failure to retrieve a prevocalic allomorph will result in use of the default allomorph prevocalically—that is, in leveling of the stem-final alternation between plain and aspirated stops, as evidenced in the appearance of innovative forms like the accusatives [mu.rɨ.bɨl] and [pu.ə.gɨl] cited above.

Because of the existence of the rule of (2b), failure to retrieve the prevocalic allomorph of (3c), on the other hand, will result in generation of an innovative *s*-final allomorph for the stem 'yut'—that is, in extension of the stem-final alternation between *t* and *s* at the expense of the earlier alternation between *t* and c^{h}, as evidenced in the appearance of innovative forms like accusative [ju.sɨl]. As we saw in section 2 with respect to umlaut in Germanic adjective paradigms, then, leveling and extension are both regularization, and result from the same mechanism, namely loss of lexical irregularity.

The analysis (2) will also account for the behavior of recent loan words ending in voiceless stops, which are often discussed in connection with the ongoing changes we have observed. In particular, before a vowel-initial clitic, stem-final *p* and *k* (*t^{h}jullip* 'tulip', *t^{h}ɨlək* 'truck') undergo only automatic voicing, but stem-final *t* undergoes the rule of (2b), so that the accusative of *int^{h}ənet* 'internet' is [in.t^{h}ə.ne.sɨl]. Unlike words like those in (3), recent loanwords have no irregular allomorphs, so their inflected forms display no variation.

Analysis (2), in particular with regard to coronal obstruents, goes back to Ko (1989) and is assumed, in whole or in part, by a variety of researchers. I thus make no claim of originality in that regard. At the same time, however, the literature contains many proposals concerning the data we have seen above that are incompatible with analysis (2). As a summary of that data in preparation for considering such proposals, key examples are repeated in (4)-(6).

(4) a. [mu.rɨ.p^{h}ɨl] ~ [mu.rɨ.bɨl] 'knee (acc.)'
    b. [pu.ə.k^{h}ɨl] ~ [pu.ə.gɨl] 'kitchen (acc.)'
    c. [pa.k'ɨl] ~ [pa.gɨl] 'outside (acc.)'

(5) a. [ju.c^{h}ɨl] ~ [ju.sɨl] 'yut (acc.)'
    b. [pa.t^{h}ɨl] ~ [pa.sɨl] 'field (acc.)'
    c. [na.ʤɨl] ~ [na.sɨl] 'day (acc.)'

(6) [int^{h}ənet]/[int^{h}ənes]

(4), first of all, exemplifies variation in prevocalic forms of labial- and velar-final stems; we omit examples of stems ending in clusters. (5) exemplifies variation in prevocalic forms of coronal-final stems, and (6) exemplifies the *t/s* alternation shown by recent coronal-final loanwords, with *t* appearing before syllable boundary and *s* before a vowel. In section 4 below, we will examine three alternative explanations for the data of (4)-(6). Our conclusion will be that

all three have serious problems and that none presents a serious challenge to the reanalysis-and-regularization (below, "R&R") account based on analysis (2) and lexical representations of the form (3) that we have sketched in this section.

4. Alternatives to the reanalysis-plus-regularization account

In discussing alternatives to the account of (4)-(6) presented in section 3.3, we will set aside the proposal that the *t/s* alternation results from a phonetically motivated change (Kim 2001), as this hypothesis provides no explanation for the fact that that alternation is limited to noun inflection. We will consider first an account on which the variation in (4)-(5) is the result of a gradual shift or reassignment of underlying representations and the alternation in (6) is the result of lexicalization of apparently *t*-final loanwords as *s*-stems. Next we will take up an account that, in accordance with (2a) above, takes the default forms of noun stems to coincide with their isolation forms, but which claims that in place of the *t*-to-*s* rule of (2b), there is a set of stochastic or probabilistic rules, their strength based on lexical frequency, changing *t* into a variety of other coronal obstruents. Finally, we will consider an Optimality-Theoretic account according to which the variation in (4)-(5) is the result of differential constraint ranking, and the alternation in (6) follows from constraint ranking as well.

4.1 UR reassignment

The basic claim of what we will call the "UR reassignment" account of the variation in (4)-(5) is that that variation is a direct reflection of variation in what URs speakers postulate for the stems in question. On this account, there has been no reanalysis of default forms as coinciding with isolation forms, there is no lexically specified irregularity, and all alternations are the result of the established rules of CN and CR. Specifically, the variation in URs that underlies the variation in prevocalic forms seen in (4)-(5) will be as in (7)-(8).

(7) a. /murɨp^{h}/ ~ /murɨp/ 'knee'
b. /puək^{h}/ ~ /puək/ 'kitchen'
c. /pak'/ ~ /pak/ 'outside'

(8) a. /juc^{h}/ ~ /jus/ 'yut'
b. /pat^{h}/ ~ /pas/ 'field'
c. /nac/ ~ /nas/ 'day'

Under the UR reassignment account of (4)-(5), as a corollary of the fact that the only rules available to account for alternations are CN and CR, coronal-final loanwords will have phonological forms ending in *s*: /int^{h}ənes/ 'internet'.

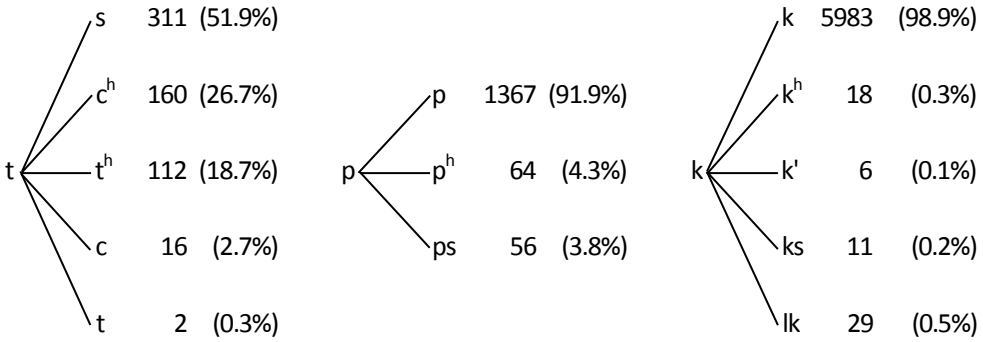
The UR reassignment account as just presented is something of an idealization, for several reasons. First, in the literature, that account is typically observed as an explanation for variation and change in coronal-final stems; labial and velar stems, which are presumably considered less interesting, are less often referred to in this connection. In some cases (Nevins and Vaux 2007:46-47), adherence to UR reassignment must be inferred from a discussion of coronal-final loanwords alone.

Finally, there are cases in which researchers appear to equivocate with regard to the choice between UR reassignment and something like the analysis of (2) above. Thus, for example Kang (2003:131) claims that "learners take the isolation form as a base" and that "a learner of Korean reanalyzes the alternation introduced by coda neutralization as a mapping from an isolation form, with a final [t], to prevocalic forms, with a final [s]." This sounds like a clear statement of (2a) and (2b) above, respectively. Immediately afterwards, however, she says "The new mapping … is extended to other nouns that are not originally /s/-final, resulting in the wholesale change of all nouns ending in a coronal obstruent to /s/-final nouns." Her conclusion, then, contrary to analysis (2), is that in the end, speakers assign *s*-final representations to all coronal-final stems, precisely as claimed by the UR reassignment account.

The most general question raised by the UR reassignment account is why the phenomenon of UR reassignment should occur at all; to my knowledge, UR reassignment of the sort postulated above is not a generally recognized change type, and the phonological literature contains no reference to a general theory of UR reassignment of which the Korean case could be seen as an example. A more specific question of the same type is why these reassignments in particular should occur. With regard to that issue, it is common to cite lexical statistics concerning the relative frequency of stem types; thus Nevins and Vaux (2007:47) note that Albright (see Albright 2008:171) finds that "56% of the morpheme-final coronals are /s/" and conclude concerning apparently *t*-final loanwords that "learners are using lexical statistics in constructing URs." Let us consider the lexical statistics of noun stem types in a little more detail.

The frequency in the Sejong Corpus (www.sejong.kr) of all obstruent-final noun-stem types is shown in (9), where in each of the three diagrams, the permitted syllable-final consonant, *t, p,* or *k,* is connected by the branches of the diagram to the segments or clusters with which that syllable-final alternates before a vowel. Data is from the concordance program Geuljabi II as applied to the .txt files of Kim and Kang (2000); the figure for stems showing the *t/s* alternation is slightly lower than that of Albright (2008:171), but the other stem counts agree closely with his, although he provides no results for clusters.

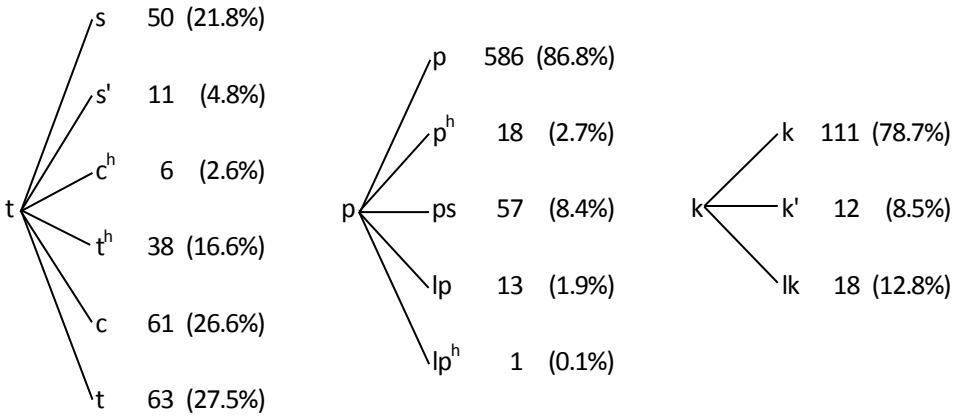
(9)



(9) makes clear that the putative target of UR reassignment, *s* for coronals, *p* for labials, and *k* for velars, is in fact the segment with the highest lexical frequency as a stem-final at each point of articulation. Could (9) therefore provide an explanation for the particular UR reassignments that appear to be observed?

That the answer to this question is negative becomes clear when we reflect that lexical imbalances like those of (9) are, in and of themselves, insufficient to produce UR reassignment. In the present case, this can be shown most concretely by citing statistics parallel to those of (9) for Korean verb stems; these are given in (10).

(10)



Apart from the fact that there is no clear winner among the coronals, the lexical statistics of (10) are broadly parallel to those of (9). But in contrast to those of (9), the statistics of (10) are phonologically inert, occasioning no apparent reassignment of URs. In verbal inflection, that is, as we have already observed, alternations due to CN and CR are

entirely stable. More generally, lexical imbalances of the sort shown in (9)–(10) are ubiquitous, but apparent reassignment of URs is relatively rare, so that it is clear that the former are not sufficient to produce the latter. We will return to the distinction between the phonologically active lexical statistics of (9) and the phonologically inert lexical statistics of (10) in section 5, where we will see that a straightforward explanation for the distinction is available in the context of the R&R account of section 3.3.

There would seem to be no convincing motivation, then, for the variable UR assignments (7)-(8), as postulated by the UR-reassignment account of (4)-(5), or for the lexicalization of apparently *t*-final loanwords as *s*-stems. There are additional problems with the UR reassignment account as well. Related to the fact that there is no general theory of UR reassignment, that account would seem to have no explanation for the diachronic profile of the associated changes, which, as we have noted, involves a long period of variation or competition between innovative and conservative forms, with the ratio of the former to the latter rising gradually as time goes on. And the UR reassignment account, by attributing the *t/s* alternation of coronal-final loanwords to their lexicalization as *s*-stems plus the low-level rule of CN, has no explanation for the productivity of that alternation, which applies exceptionlessly to turn the apparent final *t* of words like *internet* to *s* before a vowel.

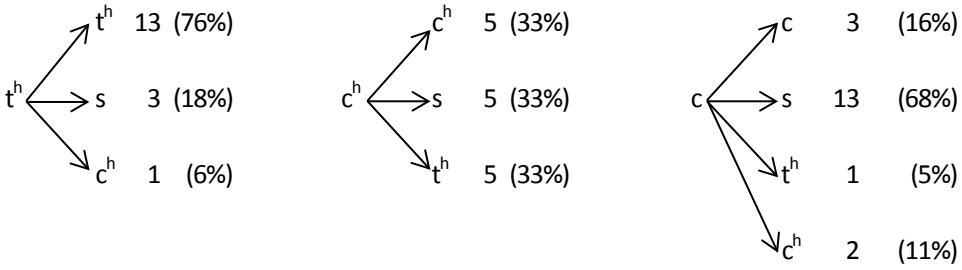
Indeed, it is not uncommon to find in the literature explicit rejections of attempts to account for productive alternations in terms of UR reassignment. Thus, with regard to English "intrusive *r*", Bermúdez-Otero (2011:2037) says, "... restructuring scenarios in which /ə/ is replaced by /əɹ/ in underlying representations ... fail to account for the regular and productive nature of *r*-intrusion ....", and with regard to Uyghur *y*-zero and *r*-zero alternations, Vaux (2008:46-47) notes that "... to say that all long-vowel roots have been historically reanalyzed as ending y or r ... relegates to the domain of arbitrary lexical content something that otherwise receives a simple phonological explanation." It seems clear, then, that there are multiple problems with the UR reassignment account of (4)-(6) and thus solid reasons to prefer the R&R account based on analysis (2) that we developed in section 3.3.

4.2 Minority tendencies and stochastic rules

The R&R account of section 3.3 makes a very clear prediction about what stems should display variation of the sort illustrated in (4)-(5), namely that a stem should show variation if and only if there is excess information—equivalently, specification of irregularity—in its lexical entry. Concretely, this means that stems ending in historical **p, *k,* and **s* should be invariant, and that the same should hold for recent loanwords ending in /p t k/. This prediction appears to be correct, as we will see below.

For stems that do have excess information in their lexical entries, of course, the R&R account predicts that conservative forms that reflect that excess information should vary with regularized innovative forms that do not. But nothing in that account precludes the appearance of additional variants that coincide with neither the conservative nor the regularized forms. The best reason we have to believe that such variants do occur is the existence of the minority tendencies, noted above, that characterize the mutual relations of stem-final coronal obstruents. In illustration of these tendencies, consider first the historical development of coronal-final stems since the Middle Korean (MK) period. The statistics in (11), from Ito (2010:363) show, for MK stems ending in **t^h, *c^h,* and **c,* how the alternations displayed by those stems changed between MK and the modern language.

(11)



For example, the middle figure of (11) shows that, of 15 MK noun stems in *c^h that have survived into the modern language, five retain the alternation t/c^h and five have switched to the dominant alternation *t/s,* but the remaining five have switched to the alternation t/t^h instead.

For the modern language, it is known that there is a complex relationship of mutual influence between *t^h-stems and *c^h-stems, with observed forms varying by dialect and conditioned by the suffixal vowel as well as by individual stems and suffixes (Kang 2005, 2007). How should we understand the apparent imposition of such a network of mutual influence on the grammatical system postulated by our R&R account?

A rough parallel to the Korean situation is arguably provided by the coexistence, in the English inflectional system, of a default procedure for forming regular verbal past tense and past participle forms (i.e. suffixation of /d/) and a network of relations of mutual influence, based on partial similarity, that characterize the past tense and past participle forms of irregular verbs (Marcus et al. 1995). If this suggestion is on the right track, the correct explanation for the minority tendencies displayed by coronal obstruents will not conflict with the R&R account, but complement it.

One recent trend in the literature on this topic, however, is the claim that the minority tendencies of coronals support the position that the phonological grammar of Korean contains, not the categorical *t*-to-*s* rule of (2b), but a set of rules taking *t* to each of the other coronal obstruents, with those rules being stochastic or probabilistic in the sense that each is associated with a reliability rating and, as a function of the reliability rating, a confidence level that specifies how likely it is to apply (Jun and Lee 2007, Jun 2010). This proposal makes no reference to the notions of irregularity and default that characterize the R&R account. It thus renders mysterious, in the absence of a theory of how the confidence levels of stochastic morphophonological rules change over time, the diachronic profile of the variation (4)-(6). A more concrete problem, however, is that the stochastic rule proposal makes what appear to be a number of incorrect predictions. I will focus first on one of those, the prediction that historical **s*-stems should display variation.

In principle, the reliability ratings of the stochastic rules taking *t* to other coronal obstruents in Korean noun stems are determined by the relative lexical frequency of the alternations they express—concretely, that is, by the statistics of the leftmost figure of (9) above. Proposals that natural language phonologies are characterized by stochastic rules based on lexical statistics (Zuraw 2000, Becker 2009) typically include a provision exempting the established vocabulary (as opposed to loanwords) from those rules on the grounds that established items have fixed, lexically listed phonological representations. In the Korean case, however, the proposal of stochastic rules is intended specifically to capture variation in the established vocabulary, so that this is not a possible move.

As a result, the stochastic rules of Jun (2010) predict that any Korean noun ending in a coronal obstruent should have a variable realization, regardless of its etymology. While stems ending in etymological *t^h, *c^h, and *c satisfy this prediction, as we have seen, the same is not true for etymological **s*-stems. Let us take a look at the evidence that the inflection of **s*-stems is essentially invariant.

A first indication in this regard is provided by the historical development of **s*-stems since Middle Korean. The statistics for **s*-stems parallel to those given in (11) for stems ending in *t^h, *c^h, and *c are that 27 of 28 (96%) have retained the *t/s* alternation, with only one having abandoned it (*ot/oc^h* 'sumac, lacquer'). A single exception to retention of the *t/s* alternation, of course, no more compromises the regular or default status of that alternation than the change from *dived* to *dove* as the past tense of *dive* in American English compromises the status of /d/ as the default past tense suffix.

A second indication that historical *s*-stems have invariant inflection can be gained from the internet. While internet search results cannot be taken as reliable indications of frequency of occurrence, search engines do function as sensitive detectors of variation. In the case of Korean noun inflection, they reveal, in particular, variation of the types in (4) and (5). This is useful because such variation makes no appearance in standard corpora, based as those are on

texts that have undergone editing and publication.

This sensitivity of search engines to variation means that when search results show zero variation for the inflection of a given stem, we can conclude with some confidence that that stem is indeed of invariant inflection. Searches (May, 2013) for three allomorph candidates (*s*-final, t^h-final, c^h-final) of ten historical **s*-stems before six vowel-initial clitics (eighteen forms per stem) yielded virtually no evidence of variation apart from a handful of occurrences of the c^h-final allomorph of **calmos* 'mistake' (two ten-thousandths of one percent of the total number for that stem).

Finally, experimental results support the idea that historical **s*-stems are invariant in their inflection. I concentrate here on production experiments; surveys based on self-reporting have been shown to be unreliable in at least some cases (see Labov 2001:194,199), and those reporting acceptability judgments in particular run the risk of registering more variation than is actually realized in the spoken language.

Jun and Lee (2007) report an experiment in which ten native speaker college students, in a sentence-completion task, were asked to inflect eight coronal-final stems (two each ending in **s, *c, *c^h,* and **t^h*), identified in writing, in the context of each of three case clitics. For **s*-stems, in sharp contrast to the other three stem-types, there was only one response out of 60 in which the stem-final consonant deviated from the historically expected value (an occurrence of c^h before Accusative *-il*). The experiments on noun inflection reported by Do (2013) are yet more naturalistic in that no written prompts were involved, and subjects were guided in their responses only by context. As shown by charts (28), (36), and (38) (Do 2013:28-34), no subject, adult or child, produced anything but etymologically expected (i.e. *s*-stem) forms for historical **s*-stems, again in sharp contrast to coronal-final stems of other types.

There is evidence from multiple sources, then, that the inflection of etymological **s*-stems is essentially invariant, disconfirming an important prediction of the stochastic rule interpretation of variation in the realization of coronal obstruents. When we note that the mean reliability rating of Jun's (2010:160) suffix-specific rules taking *t* to *s* before a vowel is .418, and that their mean confidence level is .392, it becomes clear just how far the predictions of the stochastic rule account deviate from the facts in this case. If those figures are taken literally, that is, the stochastic rule account entails the expectation that for any stem-final coronal obstruent, historical **s* included, the likelihood of its alternating with *s* prevocalically is only about 40%.

For historical **p*-stems and **k*-stems, the stochastic account predicts only limited variation because of the numerical dominance of those unmarked stem-types among stems ending in labial and velar obstruents, respectively. The gap between the predictions of the stochastic account and those of the R&R account is thus much narrower than in the case of historical **s*-stems. Nevertheless, both internet evidence of the kind referred to above and experimental evidence (Jun and Lee 2007:15, Do 2013:27,34) suggest that, in line with the predictions of the R&R account, there is essentially no tendency for historical **p*-stems and **k*-stems to show variation (in Do's results, children in the age range 5;1–7;3 represent a partial exception (Do 2013:30)).

With respect to loanwords, internet evidence again suggests invariance, in particular with respect to *t*-final stems like /int^{h}ənet/'internet'. The exception in the search result data is that the final stops of some words show a very weak tendency to display aspiration prevocalically. In some cases, this may represent influence from the aspiration that accompanies these stops when the words in question are borrowed, with epenthetic /ɨ/, as vowel-stems (e.g. /mak^{h}et^{h}ɨ/ rather than /mak^{h}et/ for English *market*).

In sum, it seems clear that the predictions of the stochastic rule account regarding what stems should show variation are not borne out by the data, notably with respect to historical **s*-stems and *t*-final loanwords. Rather, it is the prediction of the R&R account, that variation should be observed precisely for stems whose lexical entries incorporate irregularity, that appears to be confirmed.

4.3 Optimality-theoretic accounts

The last type of alternative to the R&R account of the variation in (4)-(6) that we will consider is an account couched in the terms of Optimality Theory. In sketching an OT account of that variation, I draw substantially on the analysis of Davis and Kang (2006), but for the sake of simplicity I also depart from that analysis in several ways.

Let us start our consideration of Davis and Kang's account by noting that for them, leveling in the paradigm of stems like 'knee' and 'kitchen', on the one hand, and extension of the *t*/*s* alternation in the paradigm of stems like 'yut', on the other, are due to distinct mechanisms. For labials and velars, variation between a historically expected alternation (e.g. [p]/[p^h] or [k]/[k^h]) and the null alternation, when it occurs, is due to variable ranking of a constraint that mandates faithfulness to lexical representations (IO Faith) with respect to a constraint that mandates nonalternating paradigms (Paradigm Uniformity). For coronal obstruents, on the other hand, variation between a historically expected alternation (e.g. [t]/[c^h]) and the default alternation *t*/*s* is due to variable ranking of IO Faith with respect to the constraint(s) responsible for the latter alternation.

With respect to the motivation for the *t*/*s* alternation, Davis and Kang assume that there are two relevant constraints, an anticorrespondence constraint and an antifaithfulness constraint. This division of labor appears unnecessary, however. In fact, reference to concepts such as anticorrespondence can arguably be dispensed with altogether. First, because of the CN constraint requiring that coda obstruents be unmarked—that is, negatively specified for the features [Spread Glottis], [Constricted Glottis], [Continuant], and [Delayed Release]—a syllable-final coronal obstruent will always be *t*. As a result, a further constraint to the effect that a coronal obstruent in the environment .__ $_N$] (i.e. simultaneously syllable-initial and final in a noun stem) must be *s* will suffice to account for the *t*/*s* alternation. Call the latter constraint "Syllable-initial *s*".

The constraints that a pared-down OT account will postulate are thus the following four:

- Coda Neutralization (CN)
- Input-Output Faithfulness (IOF)
- Paradigm Uniformity (PU)
- Syllable-initial *s* (SI *s*)

Following Davis and Kang (2006:10), however, I will omit the undominated constraint CN from tableaux.

Let us now verify how conservative and innovative prevocalic variants for labial/velar-final stems and coronal-final stems will be chosen under an OT analysis based on the above constraints. For that purpose, I will assume with Davis and Kang that whole paradigms (here represented by the isolation form and the Accusative) rather than individual forms are the objects of evaluation, and will adopt their historically accurate phonological representations for stems. Consider first the stem **mulɨp^h* 'knee'. In (12) and (13), a conservative paradigm showing the alternation *p*/p^h and an innovative paradigm in which that alternation has been leveled are evaluated for satisfaction of Input-Output Faithfulness and Paradigm Uniformity.

(12)

/mulɨp^h/	IOF	PU
☞ a. mulɨp, mulɨp^hɨl		*
b. mulɨp, mulɨbɨl	*!	

In (12), IOF dominates PU, so that the conservative paradigm (12a), faithful to the lexical representation, is the winning candidate. In (13), the ranking is reversed, so that the winning candidate is the innovative paradigm in which the *p*/p^h alternation has been leveled. Note that in (13b) it is necessary to assume that automatic voicing does not create a violation of Paradigm Uniformity.

(13)

/mul+p^h/	PU	IOF
a. mul+p, mul+p^h+il	*!	
☞ b. mul+p, mul+b+il		*

Concerning the range of possible candidate paradigms, we may observe that a hypothetical paradigm whose stem alternated, but according to a different pattern from that seen in (12a) and (13a)—say, a paradigm that showed the alternation *p*/*p'* rather than *p*/*p*^{h}—would violate both PU and IOF, thus incurring a violation of the most highly ranked constraint regardless of which ranking is adopted.

Consider now the stem **juc*^{h} 'yut'. In (14) and (15), a conservative paradigm showing the alternation *t*/*c*^{h} and an innovative paradigm showing the *t*/*s* alternation are evaluated for satisfaction of Input-Output Faithfulness and Syllable-initial *s*.

(14)

/juc^h/	IOF	SI *s*
☞ a. jut, juc^h+il		*
b. jut, jus+l	*!	

In (14), IOF dominates SI *s*, so that the conservative paradigm (14a), faithful to the lexical representation, is the winning candidate. In (15), the ranking is reversed, so that the winning candidate is the innovative paradigm in which the *t*/*s* alternation has been adopted.

(15)

/juc^h/	SI *s*	IOF
a. jut, juc^h+il	*!	
☞ b. jut, jus+l		*

Just as was the case with respect to (12) and (13), a hypothetical candidate paradigm whose stem alternated, but according to a pattern distinct from that mandated by the phonological representation—say, a paradigm that showed the alternation *t*/*t*^{h} rather than *t*/*c*^{h}—would violate both SI *s* and IOF, thus incurring a violation of the most highly ranked constraint regardless of which ranking is adopted. A nonalternating paradigm would also violate both SI *s* and IOF. Since it would satisfy PU, however, excluding it requires assuming that PU is outranked by one or both of those constraints.

Finally, consider a stem whose phonological representation ends in *s,* a class that for Davis and Kang (2006:15) will include apparently *t*-final loanwords like [lobot]/[lobos] 'robot' (although they omit URs for such loanwords in tableaux). For such a stem, as shown in (16a), a paradigm displaying the *t*/*s* alternation will satisfy both SI *s* and IOF, so that that paradigm will be optimal regardless of which of those constraints is ranked more highly.

(16)

/lopos/	SI *s*	IOF	PU
☞ a. lobot, lobos+l			*
b. lobot, lobod+l	*!	*	
c. lobot, loboc^h+il	*!	*	*

The nonalternating paradigm (16b), in contrast, will violate both of those constraints, although it will satisfy PU. A paradigm showing an alternation other than *t*/*s,* finally, , as in (16c), will violate all three constraints.

The OT account of the variation in (4) and (5) just sketched is simpler than that of Davis and Kang in several respects. For example, it does without any counterpart of Davis and Kang's (2006:10) anticorrespondence constraint, which mandates alternation of stem-final *t* without specifying what *t* alternates with. The simpler account nevertheless retains a number of features that arguably render it unsatisfactory, in particular by comparison to the R&R account proposed above.

First, there is the globality that is inherent in reference to the concept "paradigm" and postulation of a paradigm uniformity constraint. Such a constraint, of course, is not specific to the Davis and Kang analysis of Korean noun inflection, but the relevant fact in the present context is that nothing similar is necessary in the R&R account, where leveling results automatically from the elimination of lexical irregularity.

Second, there is the OT account's disparate treatment of leveling and extension, as contrasted with the R&R account's unified treatment of the two phenomena as lexical simplification, given the existence of the *t*-to-*s* rule of (2b). As a result of this difference, while the R&R account has an immediate explanation for the fact that leveling in noncoronals and extension of the *t*/*s* alternation in coronals are occurring simultaneously, the OT account is forced to consider that fact a coincidence.

Third, there is the OT account's failure to illuminate the diachronic profile of the variation in (4)-(5). In particular, the following questions remain unanswered—in fact, as far as I am aware, unasked:

- Why should there be variation in the relevant constraint rankings in the first place?
- Why should that variation change over time, rather than being stable?
- Why should the variation eventually disappear, for individual stems and in general?

Finally, there is the extra formal complexity entailed by the need to specify constraint rankings in the OT account. The R&R account, by contrast, appeals to no form of computational ranking or precedence. In this connection, we may note that in the R&R account, CN, the constraint that only unmarked obstruents may occur syllable-finally, plays no role in the computation of nominal inflected forms because its results have been incorporated into lexical representations.

In sum, it seems clear that there are multiple reasons to prefer the R&R account of variation and change in the Korean noun paradigm to an OT account of the same data. Indeed, it is hard to avoid the impression that the OT framework, eschewing as it does reference to the concepts of default and (ir)regularity, is intrinsically ill-suited to deal insightfully with variation and change of the sort the Korean case exemplifies.

5. Issues of explanation

5.1 A first pass: explaining speakers' choice of the R&R analysis

To this point, our discussion of Korean noun inflection has concentrated almost entirely on descriptive issues. We have been concerned, that is, with the question of what analysis of the observed data has been adopted by speakers, and in particular how the variation illustrated in (4)-(5) is to be understood. It is also important, however, to consider issues of explanation—that is, to ask whether a proposed analysis follows from plausible principles of some simplicity and generality.

For the R&R analysis, to begin with, this means explaining the two analytic decisions of (2) above, repeated below as (17).

(17) a. Default forms of noun stems coincide with their pre-consonantal and word-final allomorphs.
b. There is a rule taking *t* to *s* when it is both syllable-initial and final in a noun stem: $t \rightarrow s / . \_\_ _N]$

In fact, there are well-known and widely accepted explanations for both (17a) and (17b). The proposal that (17a) is the result of the principle that "isolation forms are liable to be recognized as basic or underlying" was articulated by Ko (1989) and has been adopted by many researchers since (see e.g. Do, Ito, and Kenstowicz 2014). The proposal that (17b) is the result of the numerical dominance in the lexicon of stems showing the *t*/*s* alternation, as documented in (9) above, goes back to Ko (1989) as well (Jun 2010:150), and has been appealed to by scholars such as Hayes (1998) and Albright (2008). The two relevant principles are roughly as stated in (18).

(18) a. There is a tendency for speakers to take isolation forms of stems, when those exist, as phonologically basic.
b. There is a tendency for speakers to take as regular that member of a set of candidate alternations that has the highest lexical frequency.

We may provisionally adopt both (18a) and (18b), although we will attempt to place them in a broader context in section 5.2 below.

Before we do that, however, there are two points that we should note concerning appeal to the lexical statistics of (9). The first is that those statistics acquire explanatory force only in the context of treating isolation forms as basic. We noted in section 4.1 that lexical imbalances like those of (9) are, in and of themselves, insufficient to trigger instability and change in patterns of alternation, illustrating with the phonologically inert lexical statistics (10) for Korean verb stems. Ordinarily, then, there is no reason to expect lexical statistics to be phonologically active in the sense of triggering reassignment of underlying representations. Taking isolation forms as basic, however, constitutes a "neutralizing" choice of underlying representations, one that targets neutralized rather than contrastive values of alternating features or segments. This analytic choice puts the alternations of underlying *t* with *s, t*^{h}, *c*^{h}, and *c* in competition with each other for the role of regular alternation, and it is at that point that lexical statistics come into play.

The second point is that there are, broadly speaking, two ways in which lexical statistics might be reflected in the phonology, given a situation in which underlying X alternates depending on lexical item with multiple Y_i. On one of them, all of the alternations X/Y_i would be transferred as rules or constraints into the phonology, each with a tendency to apply proportional to its lexical frequency. In terms of decision theory, this would be a "probability matching" response. On the other, only the most frequent alternation would be phonologized as a rule; minority alternations would be taken as irregular and thus represented in the lexicon rather than in the phonology. In decision-theoretic terms, this would be a "probability maximization" response.

Of the accounts of the variation (4)-(5) that we have considered above, the stochastic rule account, of course, exemplifies the probability matching response to a set of alternations X/Y_i, and the R&R account exemplifies the probability maximization response. We have argued on empirical grounds against the stochastic rule approach, but in the present context, it is clear both that that account is relatively closely related to its R&R rival and that it cannot be seriously faulted on explanatory grounds. The stochastic rule account, that is, does follow from principles of some simplicity and generality, setting aside the question of whether or not those principles are correct in every detail. In contrast, the other two accounts of the variation (4)-(5) we have considered, the UR reassignment account and the Optimality Theoretic account, appear much more vulnerable in explanatory terms, in that it is difficult to imagine a principled explanation, in the terms of those accounts, for why the variation in (4)-(5) should have arisen.

5.2 Generalizing the explanatory principles

Given the importance of lexical or type frequency in determining when speakers will generalize a linguistic pattern (Bybee 2001:118ff.), the reference to lexical frequency in (18b) above suggests that it might be possible to view that tendency as a special case of a more general principle. In this section, incorporating material from de Chene (2016), we will see that that is indeed possible. Let us begin by considering how the concept of type frequency could apply to the task of choosing a basic or underlying representation for an alternating stem or affix. In doing so, I will limit the scope of the inquiry to nonautomatic alternations, setting aside automatic alternations, those governed by exceptionless or surface-true constraints. The reason for this limitation is that UR choice for any phonological feature that is subject to an automatic alternation arguably results from a straightforward criterion unrelated to type frequency: URs are simply those feature-level representations all surface deviations from which can be accounted for in terms of the need to satisfy the relevant constraint (see de Chene 2014). For neutralizing alternations, in particular, these will be contrastive rather than neutralized feature values.

For the purpose of investigating UR choice for nonautomatic alternations, we may take an inflected form to be composed of a stem inserted from the lexicon and a morphosyntactic feature complex developed by the syntax. For an alternating affix, the type frequency of each of its allomorphs is naturally defined as the percentage of eligible stems with which that allomorph occurs. For example, verbal suffixes in Japanese (other than those that begin with *t* after all stems) alternate depending on whether the verb stem ends in a consonant or in a vowel. Since about 66% of Japanese verbs are consonant-stems (see de Chene 2016 and references cited there), for each verbal inflectional suffix, the type frequency of its C-stem alternant will be 66%, and that of its V-stem alternant 34%.

The type frequency of an affix allomorph is an entirely typical instance of the informal notion of type frequency, identified as that notion is with frequency in the lexicon. The same cannot be said of the idea of applying the notion of type frequency to stem allomorphy. Nevertheless, the characterization of an inflected form as a stem plus a morphosyntactic feature complex invites a definition of type frequency for stem allomorphs parallel to that we have just proposed for affix allomorphs. Under that definition, the type frequency of a stem allomorph would be the percentage of eligible morphosyntactic feature complexes with which it combines.

For example, verb stems in Portuguese (and Romance languages generally) have both stressed and unstressed alternants. Taking a conservative variety of the language as representative, there are roughly 60 morphosyntactic feature complexes with which a stem can co-occur, each corresponding to an inflected form (six person-number combinations each for nine subparadigms; two imperative and four nonfinite forms). Of these 60 forms, precisely nine are stem-stressed. For any verb stem, then, the type frequency of its stressed alternant will be 15%, and that of its unstressed alternant 85%.

In both Japanese and Portuguese, the type frequency statistics we have cited are implicated in the choice of underlying representations for alternating morphemes. Consider first the Japanese case. Across Japanese dialects, (a) C-stem suffixes and (b) V-stem suffixes that consist of the consonant *r* plus the corresponding C-stem suffix (Conclusive *-(r)u,* Provisional *-(r)eba,* Passive *-(r)are-*) are stable, whereas other V-stem suffixes are subject to replacement by innovative substitutes that do follow that pattern. This extension of the *r*/zero alternation in suffixes is naturally seen as the result of two analytic decisions: C-stem suffixes are basic, and regular V-stem suffixes are derived by a rule that inserts *r* intervocalically at verb stem boundary. As a corollary of that analysis, V-stem suffixes not derivable by *r*-Epenthesis are irregular and subject to elimination. Crucially, the choice of C-stem suffixes as basic is naturally seen as motivated by their high type frequency, roughly double that of V-stem suffixes.

Consider now the development in Portuguese of verb stem alternations dependent on stress. The seven-vowel system of Western Romance was subject to a neutralization rule raising unstressed *ɛ ɔ* to *e o*. For verbs whose stem vowel was **ɛ* or **ɔ,* this resulted in alternations depending on whether stress fell on the stem or on a suffix: stem-stressed *nɛ̲go* 'I deny', *rɔ̲go* 'I ask', but suffix-stressed *negamos* 'we deny', *rogamos* 'we ask'. Verbs with stem vowels **e* and **o,* of course, did not alternate in this manner: *pe̲sko/pe̲skamos* 'I/we fish', *po̲do/po̲damos* 'I/we prune'. In Portuguese, however, the alternation between stressed *ɛ* and *ɔ* and unstressed *e* and *o* was extended to originally nonalternating stems like *pesk-* 'to fish' and *pod-* 'to prune' (see Williams 1938:206ff.), so that in the modern language, the only verb-stem mid vowels that fail to alternate in this way are those that are exempt for phonological reasons such as nasalization, hiatus with a suffixal vowel, or adjacency to a palatal consonant (Hensey 1972:290–291, Azevedo 2005:74).

As in the case of the Japanese *r*/zero alternation, this extension of the *ɛ/e* and *ɔ/o* alternations in Portuguese can be seen as the result of two analytic decisions. First, in a neutralizing choice of URs, unstressed alternants were taken as basic, creating an arbitrary lexical distinction between stems whose *e o* alternated with *ɛ ɔ* under stress and stems whose *e o* was nonalternating. In this situation, the alternating pattern was taken as regular, resulting in a rule of Lowering under stress in verb stems (Harris 1974:75). Non-alternating stems were originally lexical exceptions to Lowering, but came to undergo the rule as the exception feature was lost from their lexical entries. Just as was the case in Japanese, UR choice in the Portuguese example may be plausibly attributed to type frequency; as we have

noted, unstressed stem alternants have a type frequency of 85%.

We have seen that type frequency, suitably defined, has a role to play in UR choice for nonautomatic alternations, both those that affect affixes and those that affect stems. In this context, determination of the regular alternation in the Korean example on the basis of lexical frequency suggests that, at a suitable level of abstraction, there may be a single frequency-based principle that determines both what alternants are basic and what alternations are regular. For either UR choice or rule choice, this principle ("Generalized Type Frequency") will choose from a well-defined candidate set the member that has the widest distribution with respect to a specified set of cooccurring elements.

Corresponding to what we have seen in the Japanese and Portuguese examples, the candidate set in the case of UR choice will be the set of all alternants of a given stem or affix, assuming that URs for nonautomatic alternations always coincide with occurring morpheme alternants; the set of cooccurring elements will be the set of applicable morphosyntactic feature matrices (for a stem alternant) or the set of eligible lexical stems (for an affix alternant). In the case of rule choice, on the other hand, the candidate set will be the maximal set of rule candidates sharing a given input configuration or structural description—for example, the set of rule candidates {$t \rightarrow s$, $t \rightarrow t^h$, $t \rightarrow c^h$, $t \rightarrow c$} in the case of Korean coronal obstruents—and the set of cooccurring elements will be the set of stems or affixes to which the rule candidates could in principle apply. With the type frequency of each candidate defined as the percentage of eligible cooccurring elements with which it occurs, Generalized Type Frequency will choose as the winning candidate the one with the highest type frequency.

We have sketched a way of integrating (18b) above into a more general criterion for the analysis of nonautomatic alternations that applies both to UR choice and to the choice of a rule. If we now turn our attention to (18a), the tendency for speakers to take isolation forms as basic, the prospects for that kind of integration do not appear as bright, at least at current levels of understanding. The choice of isolation forms as basic is often accounted for in terms of the "salience" of isolation forms (see e.g. Kang 2003:130), and type frequency as well might be considered a kind of salience, but whether those two versions of the concept can be unified, and more generally how morphophonological salience is to be defined and measured, remain topics for future research.

6. Conclusion

The Japanese and Portuguese examples of reanalysis and regularization sketched in section 5.2 have a good deal in common with the Korean case that has been our main focus of interest. Specifically, each of the three situations involves the postulation of a dedicated word-level phonological rule subsequent to a choice of underlying representations that is either frankly neutralizing (as in the Portuguese and Korean cases) or, more generally, precludes the prediction of nonbasic forms in purely phonological terms, as is the case in Japanese. In all three cases, these analytic choices will have resulted in a great deal of lexical irregularity. In this concluding section, I would like to ask what implications cases of this sort have for our understanding of how speakers choose phonological analyses, in particular for inflectional alternations, and for our conception of naturalness as it applies to phonological rules or processes.

Phonologists often assume, implicitly or explicitly, that, in accounting for alternations, speakers are aiming to maximize some a priori desirable property of the analysis as a whole. For example, the claim that speakers aim to maximize the predictability of nonbasic morpheme alternants is embodied in the feature-counting evaluation metric of classical generative phonology (Chomsky and Halle 1968), and the claim that they aim to maximize the predictability of inflected forms is taken as a starting point in the work of Albright (2002:ix), who proposes that "learners select the base form that is maximally informative, in the sense that it preserves the most contrasts, and permits accurate productive generation of as many forms of as many words as possible."

Data of the sort we have seen here suggest a rather different picture. In all three cases we have examined, that is, it would appear that UR choice has proceeded without any regard for global considerations of predictability. This

conclusion seems more or less inescapable in cases of neutralizing UR choice, and it has been argued for in some detail in the Japanese case as well (de Chene 2016). It appears, in other words, that speakers are at times driven, in defiance of considerations of predictability, to decisions about UR choice that, while they may be principled, saddle their descendants for many generations to come with considerable amounts of lexical irregularity.

If this image of lexical irregularity as a burden on speakers is accurate to any degree, we might expect to find, perhaps in the course of the acquisition process, some evidence of that fact. While the following remarks are somewhat speculative, it seems to me that the avoidance strategies for coronal alternations in nouns that are documented by Do (2013:22-49) for children acquiring Korean are suggestive in this regard.

Do found, in the experiments we referred to in section 4.2, that children go through two stages with regard to the analysis of coronal-final stems before they achieve a state that approximates that of adults. In the first of those stages, they treat stems ending in coronal obstruents as nonalternating, apart from automatic intervocalic voicing. This stage is also recorded by Kang (2003:131 (note 3)) for her child, noted as producing [ot̬in] (i.e. [od̬in]) for adult [os̬in] 'clothes (topic)'.

It is at the next, intermediate, stage that children's responses become more interesting. Children in the intermediate group, aged 5;1 to 7;3, have abandoned the nonalternation strategy of the youngest group and can be shown to have the *t*-to-*s* rule of (17b) (=(2b)). To a good first approximation, however, they will use the *t*-to-*s* rule only when the syntactic environment, a coordinate construction in Do's experiment 2, forces them to do so. Perhaps because they are aware that, for many stems, the *t*-to-*s* rule is not the whole story, their first instinct, which it is tempting to characterize as an aversion reaction, is to avoid, by any means possible, placing stems ending in a coronal obstruent in a prevocalic environment at all. To this end, they omit case clitics, they use consonant-initial Dative *-hant^{h}e,* ordinarily restricted to animate nouns, in place of the expected *-e,* they construct compounds in which the target stem appears in first position before a consonant, or they avoid the target stem altogether, typically using a sonorant-final substitute. It is tempting to see in this elaborate web of avoidance strategies evidence for the stress that we might expect to be triggered by wide-scale lexical irregularity.

We have proposed that the cases of reanalysis and regularization we have seen here suggest a particular picture of how speakers analyze nonautomatic alternations, one in which URs are chosen without any reference to global considerations of predictability. Let us now ask whether cases of this sort have anything to tell us about phonological naturalness, particularly as it relates to phonological rules or processes.

Alternations resulting more or less directly from sound change will in general involve a close phonetic relation both between the basic and the derived alternants, on the one hand, and between the derived alternant and the environment, on the other. It would be easy to conclude that both of these conditions are necessary for a phonological rule or process to be judged natural. Phonological experimentation, however, presents a somewhat different picture. With regard to input-output phonological distance, there is reason to believe that artificial rules involving minimal (single-feature) changes are indeed learned more easily than those whose input and output differ by two features or more (Skoruppa, Lambrechts, and Peperkamp 2011). Results concerning the role of a phonetically natural environment, however, are more mixed (see Do 2018:2 and references cited there), and in a review of the artificial phonology literature, Moreton and Pater (2012:711-712) suggest that "Inductive bias … facilitates faithful acquisition of simple patterns …, but is (relatively) insensitive to their phonetic motivation."

In this context, we may observe that while the three dedicated word-level rules we have considered here, the Korean *t*-to-*s* rule, Japanese *r*-Epenthesis, and Portuguese Lowering, all have environments referring to lexical category and to the concept "stem", those nonphonetic conditions do not appear to have inhibited induction of the rules in question. Further, with respect to the relation between input and output, it is noteworthy that both of the feature-changing rules, those of Korean and Portuguese, are single-feature changes, involving one degree of aperture in the case of Korean and one degree of vowel height in the case of Portuguese. Spanish Diphthongization (e/o → ie/ue), which

has been leveled far more frequently than extended, constitutes an apparent minimal pair with the fully general Portuguese Lowering rule in this respect, consistent with the idea that there is an input-output phonological distance condition on the induction of morphophonological rules.

With regard to the relative importance of the input-output relation and the output-environment relation, then, there appears to be a clear convergence between the characteristics of the dedicated word-level rules we have seen above and the results of phonological experimentation. It would seem that in addition to their inherent interest, cases of reanalysis and regularization like those we have examined here have the potential to deepen our understanding of a number of basic issues in phonology, notably the criteria for UR choice and the nature of phonological naturalness.

References

Albright, Adam (2002) *The identification of bases in morphological paradigms*. Doctoral dissertation, UCLA.

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

Azevedo, Milton M. 2005. *Portuguese.* Cambridge: Cambridge University Press.

Becker, Michael (2009). *Phonological trends in the lexicon: the role of constraints.* Doctoral dissertation, University of Massachusetts, Amherst.

Bermúdez-Otero, Ricardo (2007) Diachronic phonology. In *The Cambridge handbook of phonology,* ed. Paul de Lacy, 497–517. Cambridge: Cambridge University Press.

Bermúdez-Otero, Ricardo (2011) Cyclicity. In *The Blackwell companion to phonology,* ed. Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice, 2019–2048. Oxford: Blackwell.

Bermúdez-Otero, Ricardo (2015) Amphichronic explanation and the life cycle of phonological processes. In *The Oxford handbook of historical phonology,* ed. Patrick Honeybone and Joseph Salmons, 374–399. Oxford: Oxford University Press.

Bybee, Joan (2001) *Phonology and language use.* Cambridge: Cambridge University Press.

Campbell, A. (1959) *Old English grammar.* Oxford: Oxford University Press.

Chomsky, Noam, and Morris Halle (1968) *The sound pattern of English*. New York: Harper and Row.

Davis, Stuart, and Hyunsook Kang (2006) English loanwords and the word-final [t] problem in Korean. *Language Research* 42.2: 253–274.

de Chene, Brent (2014) The duality of phonology: automatic and nonautomatic alternations at the explanatory level. Paper presented at the workshop on nonautomatic alternations in phonology, 47th Annual Meeting of the Societas Linguistica Europaea, Poznań.

de Chene, Brent (2016) Description and explanation in morphophonology: The case of Japanese verb inflection. *Journal of East Asian Linguistics* 25:37–80.

de Chene, Brent (2018) *r*-Epenthesis and Ryukyuan. To appear in *Gengo Kenkyu.*

Do, Youngah (2013) Biased learning of phonological alternations. Doctoral dissertation, MIT.

Do, Youngah (2018) Paradigm uniformity bias in the learning of Korean verbal inflections. To appear in *Phonology.*

Do, Youngah, Chiyuki Ito, and Michael Kenstowicz (2014) The base of Korean noun paradigms: Evidence from tone. *Korean Linguistics* 16:111–143.

Garrett, Andrew (2008) Paradigmatic uniformity and markedness. In *Linguistic universals and language change,* ed. Jeff Good, 125–143. Oxford: Oxford University Press.

Hayes, Bruce (1998) On the richness of paradigms, and the insufficiency of underlying representations in accounting for them. Handout of a talk given at Stanford University.

Harris, James W. (1974) Evidence from Portuguese for the "Elsewhere Condition" in phonology. *Linguistic Inquiry* 5:61–80.

Hensey, Fritz G. (1972) Portuguese vowel alternation. In *Generative studies in Romance languages,* ed. Jean Casagrande and Bohdan Saciuk, 285–292. Rowley, MA: Newbury House.

Hyman, Larry (1976) Phonologization. In *Linguistic studies offered to Joseph Greenberg,* ed. Alphonse Juilland, A.M. Devine, and Laurence D. Stephens, 407–418. Saratoga, CA: Anma Libri.

Ito, Chiyuki (2010) Analogy and lexical restructuring in the development of nominal stem inflection from Middle to Contemporary Korean. *Journal of East Asian Linguistics* 19:357–383.

Jun, Jongho (2010) Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19:137–179.

Jun, Jongho, and Jeehyun Lee (2007) Multiple stem-final variants in Korean native nouns and loanwords. *Eoneohag* 47:159–187.

Kang, Yoonjung (2003) Sound changes affecting noun-final coronal obstruents in Korean. In William McClure, ed., *Japanese/Korean Linguistics* 12, 128-139. Stanford, CA: CSLI.

Kang, Yoonjung (2005) The emergence of the unmarked in an analogical change. Paper presented at the workshop on phonological typology, Seoul National University.

Kang, Yoonjung (2007) Frequency effects and regularization in Korean nouns. Paper presented at the workshop on variation, gradience and frequency in phonology, Stanford University.

Kim, Hyunsoon (2001) A phonetically based account of phonological stop assibilation. *Phonology* 18:81–108.

Kim, Heung-gyu and Beom-mo Kang (2000) *Frequency Analysis of Korean Morpheme and Word Usage* 1. Seoul: Institute of Korean Culture, Korea University.

Kiparsky, Paul (1972) Review of *Geschichte der indogermanischen Verbalflexion,* by Calvert Watkins. *Foundations of Language* 9:277–286.

Ko, Kwang-Mo (1989) Explaining the noun-final change t > s in Korean. *Eoneohag* 11:3–22.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors.* Oxford: Blackwell.

Marcus, Gary F., Ursula Brinkmann, Harald Clahsen, Richard Wiese, and Steven Pinker (1995) German inflection: the exception that proves the rule. *Cognitive Psychology* 29:189-256.

Martin, Samuel (1992) *A Reference Grammar of Korean*. Tokyo: Tuttle.

Moreton, Elliott, and Joe Pater (2012) Structure and substance in artificial-phonology learning, part II: substance. *Language and Linguistics Compass* 6:702–718.

Nevins, Andrew, and Bert Vaux (2007) Underlying representations that do not minimize grammatical violations. In *Freedom of Analysis?*, ed. Sylvia Blaho, Patrik Bye, and Martin Krämer, 35–61. Berlin: Mouton de Gruyter.

Ringe, Don (2006) *From Proto-Indo-European to Proto-Germanic.* Oxford: Oxford University Press.

Ringe, Don, and Ann Taylor (2014) *The development of Old English.* Oxford: Oxford University Press.

Skoruppa, Katrin, Anna Lambrechts and Sharon Peperkamp (2011) The role of phonetic distance in the acquisition of phonological alternations. In Suzi Lima, Kevin Mullin, and Brian Smith, eds. *Proceedings of the 39th Annual Meeting of the North Eastern Linguistic Society,* 717-729. Amherst, MA: Graduate Student Linguistic Association, University of Massachusetts, Amherst.

Sohn, Ho-Min (1999) *The Korean language.* Cambridge: Cambridge University Press.

Vaux, Bert. 2008. Why the phonological component must be serial and rule-based. In *Rules, constraints, and phonological phenomena,* ed. Bert Vaux and Andrew Nevins, 20–60. Oxford: Oxford University Press.

Williams, Edwin B. (1938) *From Latin to Portuguese.* Philadelphia: University of Pennsylvania Press.

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

dechene@waseda.jp