Title: Exceptionality and derived-environment effects: A comparison of Korean and Turkish

Author: Adam J. Chong

Affiliation: Queen Mary University of London

Corresponding Address:

Adam J. Chong

Department of Linguistics

School of Languages, Linguistics and Film

Queen Mary University of London

Mile End Road

E1 4NS London

United Kingdom

Tel: +44 20 7882 6638

Email: a.chong@qmul.ac.uk

Keywords: derived-environment effects, phonotactics, exceptionality, gradience, Korean, Turkish

Acknowledgements:

This work has benefited from discussion and feedback from Ricardo Bermúdez-Otero Robert Daland, Bruce Hayes, Sharon Inkelas, Karen Jesney, Sharon Peperkamp, Stephanie Shih, Brian Smith, Megha Sundara, Kie Zuraw and the audiences at the University of Melbourne, UC Berkeley, UCLA, U. Hawaii, NYU, QMUL, the University of Manchester, SCaMP 2016 at UCSD and AMP 2016 at USC. I am grateful to Joo Hee Oom for help processing parts of the NAKL corpus. Finally, I would also like to thank three anonymous reviewers, an associate editor, and the editors - Ellen Kaisse and Colin Ewen - at *Phonology*, whose comments and suggestions on earlier versions of this manuscript have greatly improved it. An earlier version of this work appeared as a chapter in my 2017 dissertation. This work was funded by a UCLA Dissertation Year Fellowship. All remaining faults are my own.

Abstract:

Morphologically-derived environment effect patterns (MDEEs) are well-known examples where static phonotactic patterns in the lexicon mismatch with what is allowed at morphological boundaries - alternations. Analyses of MDEEs (e.g. Kiparsky, 1993; Lubowicz, 2002) generally assume that the alternation is morphologically general and the static phonotactic patterns are productive. That is, that the sequences repaired across a morpheme boundary are phonotactically well-formed. This paper presents the results of corpus studies and phonotactic modelling simulations examining the phonotactic patterns in the lexicon of languages with MDEEs, focusing on two well-known cases: Korean palatalisation and Turkish velar deletion. I show that, in Korean, sequences which are repaired at the morpheme boundary through palatalisation are under-attested in the lexicon. A computational learner is able to learn a markendess constraint that drives palatalisation, suggesting that the pattern is one of exceptional non-undergoers. This contrasts with Turkish where the relevant constraint motivating velar deletion at the morpheme boundary is unavailable from pure phonotactic learning, and where the alternation is an example of exceptional triggering. These results indicate that MDEEs are not a unitary phenomenon, highlighting the need to examine these patterns in closer quantitative detail.

(189 words)

1. Introduction

Phonological alternations at morphological boundaries often reflect morpheme-internal static phonotactic patterns (Chomsky & Halle, 1968; Kenstowicz & Kisseberth, 1977; McCarthy, 2002). For example, in Kirundi (Meeusen, 1959; Rodegem, 1970; data from Kenstowicz and Kisseberth 1977, 1979), vowels preceding nasal-consonant (NC) clusters are [+long] as shown in (1):

(1) Vowels are [+long] before NC clusters within stems

a.	[umu-ruːndi]	'a Rundi person'	*[umu-rundi]
b.	[ku-geːnd-a]	'to go'	*[ku-gend-a]

At the same time, this static phonotactic generalisation is also enforced across morpheme boundaries by a phonological alternation, vowel lengthening. Underlying short vowels in the prefixes /ku-/, /ba-/ and /umu-/ lengthen when prefixed before a stem containing an initial NC cluster as in (2a), (2c) and (2e) but not a singleton consonant in (2b), (2d) and (2f). Thus, both the tautomorphemic static phonotactic generalisation and the heteromorphemic dynamic generalisation involving the phonological alternation can be captured using the same rule or constraint.

(2) Vowels lengthen before NC clusters across morpheme boundaries

a.	/ku-n-dor-a/	→	[kuːndora]	'to look at me'
b.	cf. /ku-ror-a/	→	[kurora]	'to look at'

c. /ba-n-taba:re/ → [baːntabaːre] 'that they help me'

d. cf. /ba-taba:re/ → [bataba:re] 'that they help'

e. /umu-ntu/ → [umuːntu] 'person'

f. cf. /umu-gabo/ → [umugabo] '(married) man'

These two types of generalisations (static generalisations about the lexicon and dynamic generalisations about phonological alternations) do not, however, always go hand-in-hand. Morphologically derived environment effects (MDEEs; also known as non-derived environment blocking; e.g. Kiparsky, 1973, 1993) are one such example of this mismatch (see Paster (2013) for a recent review of other examples). A textbook example of an MDEE is Korean palatalisation. At morpheme boundaries, underlying stem-final coronal stops /t/ and /t^h/ palatalize to [c] and [c^h] respectively before a suffix-initial [i] as in (3) (Kiparsky, 1973; Iverson & Wheeler, 1988; Kiparsky, 1993; Cho, 2001, *a.o.*).

(3) Palatalisation across morpheme boundaries: /t, t^h/ → /c, c^h/ before /i/ and /j/[1]

a. /ma**t**-i/ → [ma**c**i] 'eldest-NOM'

b. /pa**t^h**-i/ → [pa**c**^{h}i] 'field-NOM'

c. /pa**t-h**jʌ-jo/ → [pa**c^h**ʌjo][2] 'is butted'

[1] The tense stop, /t*/, does not occur word-finally (e.g. Sohn, 1999). In this paper, I transcribe the palatal consonants using the symbol for the palatal stop, although these are often transcribed using the symbol for the alveolo-palatal affricate /tɕ/.
[2] An independent process ensures that the lax stop /t/ followed by /h/ becomes aspirated and that the onglide in /jʌ/ deletes post-consonantally.

Palatalisation, however, fails to apply when the target consonant (/t/ or /t^h/) and trigger (/i/ or /j/) are within the stem (i.e. tautomorphemic) in both native words and loanwords. Thus /ti/ and /t^hi/ sequences which are repaired at the morpheme boundary are nonetheless attested within stems where they surface faithfully as in (4).

(4) Blocking of palatalisation tautomorphemically:

a. /mati/ → [mati] 'knot, joint'

b. /t^him/ → [t^him] 'team'

Patterns such as these have continued to pose a challenge for phonological theory, starting with Kiparsky (1973) (for a recent review and proposal see Inkelas, 2015). Previous analyses of such patterns in rule-based models (Chomsky & Halle, 1968) or in constraint-based models like Optimality Theory (Prince & Smolensky, 1993/2004) have focused primarily on protecting morpheme internal non-derived sequences (such as /ti/ in Korean) while ensuring that the very same sequences always alternate if they occur due to morpheme concatenation, i.e. the derived-environment condition (5).

(5) DERIVED-ENVIRONMENT CONDITION:

Morphological derivedness is a necessary and *sufficient* condition for a process to occur(variously stated as the Strict Cycle Condition or the Revised Alternation condition; e.g. Kiparsky, 1982, 1993)

This has been achieved through a number of theoretical tools such as underspecification (Kiparsky 1993), interleaving morphological operations and phonological ones (Wolf, 2008), conjoined constraints (Łubowicz, 2002), and reference to new or old input (Comparative Markedness; McCarthy, 2003), indexed constraints (Pater, 2007), amongst others.

A number of authors (e.g. Anttila, 2006; Hammond, 1992; Inkelas, 2011), however, have argued that what are often seen as canonical cases of MDEEs do not satisfy (5), insofar as a derived environment (i.e. where the target and environment of a rule are from two different morphemes) does not actually guarantee that a particular process would apply. A derived environment, while a necessary condition for a particular process to apply, is by no means a *sufficient* condition.

One such case that has been examined in greater detail is Finnish assibilation (Anttila, 2006; cf. Hammond, 1992; Kiparsky, 1973, 1993). Stem-final /t/s in Finnish become [s] before /i/. This rule is generally characterized as only occurring across a morpheme boundary as in (6); it fails to apply within stems (7).

(6) /t/ → [s] /__i across a morpheme boundary (*ti):

a.	/halut-i/	→	[halusi]	‘want-PAST’
b.	/hakkat-i/	→	[hakkasi]	‘beat-PAST’
c.	cf./halut-a/	→	[haluta]	‘want-INF’
d.	cf. /hakkat-a/	→	[hakkata]	‘beat-INF’

(7) /ti/ sequences surface faithfully within stems:

e. /tilat-i/ → [tilasi] 'order-PAST' *[silasi]

f. /koti/ → [koti] 'home' *[kosi]

Yet the reality in the data is more nuanced. Anttila (2006), citing Karlsson (1983), shows that not all /i/-initial suffixes actually trigger assibilation. Assibilation only occurs uniformly with the three suffixes shown in (8). Many other /i/-initial stem-level suffixes fail to trigger assibilation despite satisfying the phonological (and morphologically-derived) environment for process application as seen in (9) (Kiparsky, 2003). In at least one case, the suffix variably triggers assibilation as in (10). I refer the reader to Anttila (2006) for a full analysis of these patterns (all data in 8-10 are taken from Anttila, 2006: 900-901).

(8) Triggering suffixes.

a. Plural /-i/: /vuote-i-nA/ → *vuosina* 'year-PL-ESS'

b. Past tense /-i/: /huuta-i-vAt-kO/ → *huusivatko* 'shout-PST-3P.PL-QUE'

c. Superlative /-impA/: /uute-impA-nA/ → *uusimpana* 'new-SUP-ESS'

(9) Non-triggering suffixes.

a. Instrumental /-ime/: /lentä-ime-n/ → *lentimen* 'fly-INST-GEN' (**lensimen*)

b. Conditional /-isi/: /tunte-isi/ → *tuntisi* 'feel-COND' (**tunsisi*)

(10) Variable trigger.

a. Adjectival derivational suffix /-inen/ → /vete-inen/ → *vesinen~vetinen* 'watery'

The data from Finnish suggests that an account of its putative MDEE pattern cannot simply appeal to the derived environment condition. While assibilation does indeed only occur in derived-environments, as expected under the traditional notion of MDEEs, it certainly does not occur in all possible derived environments where the phonological conditions are met. Even when it does occur, it is not entirely categorical. These patterns violate the strong interpretation of (5). The generalisation about assibiliation then seems to be about the specific triggering morphemes, rather than derived environments more generally.

Apart from the derived-environment condition, analyses that aim to capture the fact that tautomorphemic /ti/ sequences surface faithfully in Finnish, for example, also assume that such sequences are completely phonotactically well-formed (11).

(11) PHONOTACTIC WELL-FORMEDNESS:

Morpheme-internal sequences are phonotactically well-formed.

While having even just one lexical item is sufficient for a particular sound sequence to be attested, phonotactic knowledge, all else being equal, seems not to be purely sensitive to the presence or absence of a particular sequence, but rather to the frequency with which those sequences occur. A large body of evidence has shown that speakers possess gradient well-formedness intuitions based on the statistical properties of attested sequences in the lexicon (e.g., Bailey & Hahn, 2001; Coetzee & Pater, 2008; Coleman & Pierrehumbert, 1997; Frisch, Pierrehumbert, & Broe, 2004; Frisch & Zawaydeh, 2001;

Hay, Pierrehumbert, & Beckman, 2003; Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000; Albright 2009; Daland et al. 2011; Hayes & White 2013) and that listeners also use gradient well-formedness constraints in speech processing (Frisch, Large, & Pisoni, 2000; Kager & Shatzman, 2007).

With this in mind, what previous analyses of MDEEs ignore is the extent to which sequences repaired heteromorphemically are attested in the lexicon, and whether these sequences are entirely phonotactically well-formed. For example, Łubowicz (2002), examining Polish velar palatalisation, concedes that the protected stem-internal sequences cited all appear in loanwords. She argues, however, that these words have been wholly incorporated into the native grammar (Rubach, 1984) since palatalisation applies in these words across the morpheme boundary, arguing against these examples being purely exceptions to the phonological rule. Yet this argument ignores the question of how frequent and, consequently, phonotactically well-formed such protected sequences actually are in the lexicon. In the same way that these patterns are not as general as was previously thought in derived environments, as is the case for Finnish, it is possible that the static (non-derived) patterns in the lexicon are not as widespread or well-formed as is traditionally assume. Taking Korean as an example, this would entail that not only is a sequence [ti] repaired at a morpheme boundary (due to a constraint like *ti), but a computational learner, and by implication native speakers, might learn some dispreference for such sequences even when they occur within stems.

In this paper, I address this possibility by examining and comparing whether the morpheme-internal sequences involved in the alternations in derived environments are truly grammatical in two well-known MDEE patterns: Korean palatalisation and Turkish velar deletion. I leave an experimental investigation of native speaker knowledge for future work. In §2 and 3, I present the results of corpus studies and computational phonotactic learning simulations of both Korean and Turkish. In each case, I also provide historical background regarding the MDEE patterns in each language. Then in §4, I discuss how these patterns should be accounted for using indexed constraints. The implications of these results are taken up in §5, where I argue that various examples of MDEEs, while structurally similar superficially, are actually quite different from each other, highlighting the need to examine these cases with closer quantitative scrutiny.

To preview the results, non-palatalised sequences in Korean are under-represented in the lexicon and a computational learner learns a phonotactic constraint that penalizes such sequences. In Turkish, however, intervocalic velars are not sufficiently under-represented in the lexicon to be penalized by a computational phonotactic learner. Thus, the crucial sequences involved in the alternation across morphemes are phonotactically well-formed in Turkish, but not in Korean. Two putatively similar cases of MDEEs exhibit different morpheme-internal phonotactic generalisations, providing evidence against the analysis of these patterns in a unified way. Instead, Korean palatalisation, is an example of exceptional blocking (non-undergoers) of a more general phonotactic constraint, whereas Turkish velar deletion, is an example of exceptional triggering specific to individual suffixes.

2. Korean palatalisation

2.1. Further background and historical origins

How did the current mismatch between alternations and phonotactics, described in §1 develop historically in Korean? The origin of the current ostensibly derived condition on palatalisation dates back to Early Modern Korean (circa early nineteenth century; Y. Y. Cho, 2009; Lee & Ramsey, 2011 and references therein). At the start of the 19th century, palatalisation was an obligatory process that was an across-the-board sound change that neutralized the coronal stops /t, t^h, t*/ to their corresponding palatal affricate counterparts /c, c^h, c*/ before the high front vowel /i/ and the palatal glide /j/. This was a process that applied both within and across morpheme boundaries. Thus, for a time both the static phonotactic generalisation about the lexicon as well as the dynamic generalisation motivating alternations were in agreement.

The current mismatch in generalisations has three sources (examples are from Cho, 2009; Lee & Ramsey, 2011). The first was the monophthongisation of /ɨ̯i/ sequences to [i] which occurred following the sound change that involved palatalisation. Consequently, words that had underlying /tɨ̯i/ or /t^hɨ̯i/ became previously unattested [ti] and [t^hi] (12).

(12) Source 1: historical monophthongisation of /ɨ̯i/.

a. *ʌtɨ̯i > ʌti 'where'

b. *matɨ̯i > mati 'joint'

The fact that surface [ti] or [t^hi] were not palatalised represents an example of diachronic counterfeeding opacity. In principle, the monophthongisation of /ɨ̯i/ could have fed the palatalisation process, but this did not occur, resulting in a generalisation that was not entirely surface true. This same process is also reflected in synchronic monophthongisation of /ɨ̯i/ sequences that were the result of morpheme concatenation (13).

(13) Source 2: synchronic monophthongisation of /ɨ̯i/:

a. /t*ɨ-ita/ → [t*ɨita] ~ [t*ita] 'to become aware'

eye-PASS-PRED

b. /tʰɨ-ita/ → [tʰɨita] ~ [tʰita] 'to be open'

open-PASS-PRED

The final source of [ti] and [tʰi] sequences are loanwords borrowed from English and other European languages which were systematically borrowed in faithfully. Some examples are given in (14).

(14) Source 3: Loanwords from English

a. /sitilom/ from 'CD-ROM'

b. /antʰikʰɨ/ from 'antique'

Cho's (2009) broader observation in relation to MDEEs is the fact that many putative cases have a similar historical origin to that outlined for Korean: a particular phonological process had historically applied across-the-board, both tautomorphemically and heteromorphemically, but sequences which were previously unattested were reintroduced through borrowing into the language as well as other independent phonological processes. She points out that this is the case with other well-known examples of MDEEs such as Chamorro Vowel Lowering (Chung, 1983), Finnish Vowel Coalescence (Anttila, 2009) and Polish First Velar Palatalisation (Łubowicz, 2002). Cho (2009) further suggests that words that are exceptions to the more general palatalisation rule are marginal in the lexicon because of their historical origins. In what follows, I examine Cho's claim about the marginality of [ti] and [t^hi] sequences in more detail in a corpus study, before turning to its consequences for phonotactic learning in a computational learning simulation.

2.2. Corpus study: NAKL

In order to investigate the lexical trends pertaining to [ti] and [t^hi] sequences in Korean, I examined a corpus compiled by the National Academy of Korean Language[3] (NAKL, 2003). The NAKL corpus contains over 50,000 frequently used Korean words (native and Sino-Korean), including loanwords, as well as corresponding frequencies of each word from various, usually print, sources. Some of these words contain derivational suffix boundaries. To facilitate analysis, the corpus was first pre-processed and each syllable block was split up into its component Korean letters or digraphs using the grapheme-to-phonetic conversion system of Kim, Lee and Lee (2002). The system also applies regular *neutralizing* phonological rules at the appropriate morphological boundaries detected in the process of conversion. The corpus contains lexemes and thus any morphological

[3] Now the National Institute of Korean Language

boundaries would be due to derivational suffixes.[4] Note that the conversion system only applies transformations if there is already an orthographic character available for the resulting sound, thus this does not reflect any purely allophonic changes (such as intervocalic voicing).

2.3. Results

After excluding any duplicate items in the corpus, there were a total of 53,196 lexical items. The number of words in the corpus contained the consonant [t], [t^h] or [t*] followed by [i] or [j] was first calculated (I will refer to these as [Ti] or [Tj] respectively, and collectively as [TI]). Although there is no overt evidence of palatalisation of /t*/ because these do not occur word-finally, I treat the coronal stop series here as a natural class, since these all participated in the historical sound change. A count of [TI] entries in the corpus is given in Table 1. Out of a total of 53,196 words in the corpus, only 436 contained [TI] sequences, less than 1% of the lexicon. Further, out of these words, 284 (~65%) are loanwords (e.g. /t^him/ = team, etc.). These sequences, therefore, are rare in terms of absolute type frequency in the corpus.

[4] An analysis that uses a smaller corpus (~5000) also from NAKL – the "word list for Korean learners" (Cho, 2003) – that contains only roots with no forms with derivational morphology was also conducted. A similar quantitative pattern was obtained. The result of the analysis of the larger corpus is presented here to facilitate investigation into different lexical strata.

Table 1. No. of words that contain [ti], [t^hi], [t*i], [tj], [t^hj] and [t*j] in NAKL corpus (and by lexical strata).

CV type	Entire Lexicon	Native	Sino-Korean	Loanword
[ti]	208	68	5	135
[t^hi]	167	30	4	133
[t*i]	32	28	4	0
[tj]	14	5	0	9
[t^hj]	15	4	4	7
[t*j]	0	0	0	0
Total	**436**	**135**	**17**	**284**

It is possible, though, that the rarity of words with such sequences is merely attributable to either the overall rarity of the coronal stop series or the high front vocoids. To ascertain the extent of the rarity of [TI] sequences given the independent frequency of its components segments, a two-by-two contingency table was constructed (Table 2) that compared the frequency of occurrence of these sequences compared to other CV combinations. Observed/Expected (O/E) values were calculated for each cell (e.g. Coetzee & Pater, 2008; Frisch & Zawaydeh, 2001). 'Observed' (O) values are the total number of sequences of each CV combination found in the corpus. 'Expected' (E) values are how frequently each CV combination is expected if each C and V co-occurred based on chance.[5] An O/E value of 1 indicates that a particular sequence occurs at the expected rate of occurrence. O/E values above 1 indicate over-representation and O/E values under

[5] Expected values were calculated by taking the product of the relevant marginal totals (row and column) and dividing it by the grand total.

1 indicate under-representation. The percentages in bold (column %) indicate the proportion of vowels found in a particular consonantal context and the percentages in italics (row %) indicate the proportion of consonants in a particular vocalic context. The marginal percentages (bottom row and final column) indicate the expected proportions.

While Wilson & Obdeyn (2009) argue against the use of O/Es for examining lexical statistics, these are presented here to illustrate the general statistical pattern. Furthermore, since chi-square tests which are often used to analyse contingency tables are sensitive to sample sizes where small differences can be significant if the sample is large enough (Lantz, 2013), I do not report any *p*-values here. Instead, the O/E values described below serve primarily to describe the relative underattestation in the data, with more robust evidence to be presented later using a computational phonotactic learner.

Table 2. Occurrence of CV combinations: by consonant type (T vs. other Cs) and vowel type (i, j vs. other Vs). Expected counts are in parentheses. Percentages in bold: row percentages; Percentages in italics: column percentages.

	[i, jV]	other Vs	Expected % of Cs
[t, t^h, t*]	454 (5,805) **1.63%** / *1.43%* O/E = 0.08	27,424 (22,073) **98.37%** / *22.75%* O/E = 1.24	*18.31%*
other Cs	31,247 (25,896) **25.12%** / *98.57%* O/E = 1.21	93,112 (98,463) **74.87%** / *77.25%* O/E = 0.95	*81.69%*
Expected % of Vs	**20.82%**	**79.18%**	

What is clear from Table 2 is that the actual observed number of [TI] sequences in the NAKL corpus is about a tenth of what would be expected due to chance (randomly combining each C and V). While we expect about 18 % of consonants to be [T], only 1.43% of Cs in the [I] context are [T], indicated by the percentages in italics. Similarly, while we expect about 21% of vowels to be [I], only 1.63% of [I]s occur with [T] (the percentages in bold**)**. [TI] sequences occur at about a tenth the rate that we would expect them to occur given the independent occurrence of [T] and [I].

Table 3. Occurrence of CV combinations: by consonant type (CH vs. other Cs) and vowel type (i, j vs. other Vs).

	[i, jV]	other Vs	Expected % of Cs
[c, c^h, c*]	5,944 (5,501) **22.50%**/*18.75%* O/E = 1.08	20,473 (20,916) **77.50%**/*16.98%* O/E = 0.98	*17.35%*
other Cs	25,757 (26,200) **20.47**/*81.25%* O/E = 0.98	100,063 (99,620) **79.53%**/*83.02%* O/E = 1.00	*82.65%*
Expected % of Vs	**20.82%**	**79.18%**	

As a comparison, Table 3 shows the same calculations for [c, c^h, c*] and [i] or [jV] sequences ([CHI] collectively). In this case, we see instead a small over-representation of [CHI] sequences in the corpus. This is perhaps expected given that historically /TI/ sequences palatalized to [CHI] across-the-board, as was discussed above.

When each of the three strata of the Korean lexicon is examined in more detail (native, Sino-Korean and loanword), a similar statistical pattern across all of them is observed. These statistical trends are further supported in a corpus of Child-Directed Speech (Ryu, 2012). Thus, the corpus investigation so far supports the hypothesis that [TI] sequences, while attested in the Korean lexicon, are actually exceedingly rare and marginal, with a majority of such words being loanwords (see Chong, 2017 for full details).

In the next section, I present the results of a computational learning simulation aimed at ascertaining whether there is sufficient evidence for a computational learner to learn a markedness constraint penalising [TI] sequences in Korean.

2.4. Learning a phonotactic grammar of Korean

In the previous section, I presented evidence that [TI] sequences were numerically rare in the Korean lexicon. In this section, we confirm this qualitative observation by examining if a computational learner – the UCLA Phonotactic Learner (Hayes & Wilson, 2008) – assigns a penalty to these sequences in a phonotactic grammar. The learning data for the simulation is the entire NAKL corpus, including loanwords, analysed in the previous section. The motivation for including loanwords here was due to the fact that we are assuming that in an early stage of phonotactic learning, a child does not have explicit knowledge of lexical strata, thus as far as they are concerned there is no difference between a native, Sino-Korean word or loanword.

The learner was fed the feature system modified from Cho (2012) (see Table A1 in Appendix A). Following both Hayes & Wilson (2008) and Cho (2012), we used both privative and contrastive feature underspecification to control for the total number of natural classes. For the three-way laryngeal contrast, only tense stops are assigned [+constricted glottis][6], while aspirated stops are assigned [+aspirated]. Thus, lax stops are

[6] Cho (2012) uses the feature [+tense] instead. Note that Cho (2012) ultimately claims that both tense and aspirated stops should be specified as [+tense]. The model output does not change using this feature specification. Cho (2012) also includes some allophones that we have not included in the current simulation since the transcription used in the simulations are only 'semi'-allophonic.

[-aspirated, -const. glot] (Cho, 2012). The feature [+spread glottis] is only specified for /h/. As for the vowels, I have adopted a seven-vowel system /i, e, a, ʌ, o, ɨ, u/. As /e/ and /ɛ/ are now merged in most speakers' productions (Eychenne & Jang, 2015; Shin, Kiaer, & Cha, 2013), these categories were both collapsed to /e/. Glide-vowel sequences (diphthongs) were assumed to be two separate segments for the purposes of this simulation: five /j/-diphthongs (/ja, jʌ, jo, ju, je), four /w/-dipthongs (/wa, wʌ, we, wi/) and one /ɨ/-diphthong (/ɨi/). Since /wɛ/ is pronounced as /we/, these were all replaced with /we/. The vowels /ø/ and /y/ were coded as /we/ and /wi/ respectively, following Cho (2012).

The phonotactic learner was specified to only find bigram constraints, with a maximum number of constraints set at 180, although the result does not change if it is specified to find trigram constraints. The learner was initialized with the constraint *[-strident][-high,-back] (i.e. *[TI]) and we were interested if this constraint was assigned any weight. The O/E accuracy threshold for constraints was set at 0.30, following the simulations done by Hayes & Wilson (2008). Since type frequency is typically implicated in the learning of phonotactic constraints over the lexicon (e.g. Pierrehumbert, 2003; Richtsmeier, 2011), each input had a frequency of 1. All other parameters were set at default.

2.5. Results

As a first check of the results of the simulations, the highest weighted constraints were inspected to see if these corresponded to well-known phonotactic constraints in Korean.

The top 10 constraints are shown in Table 4. These conform to well-known phonotactic restrictions in Korean. For example, [ŋ] can only occur in syllable-final position (6 and 8 in Table 4). Thus, unsurprisingly, the learner assigns a penalty score of 6.56 to the nonsense form /ŋam/, indicating that this is not phonotactically well-formed.

Table 4. Top weighted constraints learned from NAKL corpus

No.	Constraint	Weight	Description
1.	*[-consonantal,-syllabic][-syllabic]	7.667	No glides before another consonant
2.	*[-consonantal,-syllabic][+word_boundary]	7.418	No glides word-finally
3.	*[+aspirated][+consonantal]	7.2	No aspirated stops before a consonant
4.	*[-sonorant][+consonantal,+sonorant]	7.159	No obstruents before a sonorant
5.	*[+aspirated][+word_boundary]	7.12	No aspirated stops word-finally
6.	*[+word_boundary][+sonorant,+dorsal]	6.656	No word-initial [ŋ]
7.	*[+const_glot][+consonantal]	6.516	No tense stops before a consonant
8.	*[-syllabic][+sonorant,+dorsal]	6.465	[ŋ] cannot occur following consonants
9.	*[+nasal][+lateral]	5.948	*nl
10.	*[-back,-syllabic][+high,-round]	5.746	*ji, *jɨ

Crucially, the learner assigns a penalty score of 1.861 to words like [mati] and [t^him] but a perfect score of 0 to comparable palatalised forms [maci] and [c^him]. That is, while this is clearly not a categorical constraint, the learner nonetheless assigns a nontrivial penalty to forms with [TI] sequences. As a point of calibration, the harmony penalty again [TI] sequences is in the range of the constraint against [θ] and [ð] or [tw] onsets in English (Hayes & Wilson, 2008: 399) which indicates that these sequences are rare and that constraint is violable. In the current case, the ultimate penalty score arises from the pre-specified constraint (weight = 1.492), in addition to another constraint, [-const_glot][+high,-back] (weight = 0.369), that specifically penalises [ti] and [t^hi] sequences individually. The fact that these were induced in addition to the pre-specified constraints provides some further evidence that such sequences are dispreferred.

Using default parameters, we have shown how a probabilistic phonotactic learning model assigns a sizable weight to a constraint penalizing [TI] sequences despite these forms actually existing in the lexicon of Korean. We can conclude, therefore, that the statistical under-representation we found does indeed translate into a *well-formedness penalty* for words with [TI]. This suggests that the implicit assumption in analyses of MDEEs that such sequences occur freely in the lexicon is empirically not supported. In the next section, I examine another well-known example of MDEEs: Turkish velar deletion. I show that the surface similarity in patterns belies some crucial differences.

3. Case study II: Turkish velar deletion

3.1. Introduction and historical background

Turkish velar deletion (Inkelas, 2000; Inkelas & Orgun, 1995; Inkelas, Orgun, & Zoll, 1997; Lees, 1961; Lewis, 1967; Sezer, 1981; Zimmer & Abbott, 1978) is another oft-cited case of an MDEE. Velars delete intervocalically if at a morpheme boundary (15), but are protected from deletion within morphemes (16) (data from Inkelas, 2011, 2015):

(15) Suffix-boundary deletion

a.	/bebek-In/[7]	→	[bebein]	'baby-GEN'
b.	cf. /bebek/	→	[bebek]	'baby-NOM'
c.	/ipek-A/	→	[ipee]	'cotton-DAT'
d.	cf. /ipek/	→	[ipek]	'cotton-NOM'
e.	/arkeolog-I/	→	[arkeolou]	'archeologist-ACC'
f.	cf. /arkeolog/	→	[arkeolog]	'archeologist-NOM'

(16) Deletion blocked morpheme-internally (from Inkelas 2011, 2015)

a.	/hareket/	→	[hareket]	'motion'
b.	/sigorta/	→	[sigorta]	'insurance'
c.	/sokak-A/	→	[sokaa]	'street-DAT'
d.	/mekik-A/	→	[mekie]	'(weaver's) shuttle'

Inkelas (2011, 2015; see also Sezer, 1981; Pycha, 2008 for further discussion) points out that while velar deletion occurs in both native words and loanwords as well as in morphologically simplex and complex stems, it fails to apply to verb roots, although the phonological conditions are met as in (17), a minimal pair /gerek/ which can either be a noun or verb. While the noun undergoes deletion (17a), the verb does not (17b) despite satisfying the phonological conditions for velar deletion to occur.

(17) Verbal roots

a. /gerek-Ijor/ → [gerekijor] 'is necessary-PROG'

[7] Capital vowels indicate vowels that undergo vowel harmony.

b. cf. /gerek-i/ → [gerei] ‘need-ACC’

Furthermore, deletion does not seem to apply when the velar consonant is suffix-initial as in (18) compared to when it is stem-final as seen in (15) despite both contexts being morphologically derived.

(18) /-ki/ suffix

a. /sene-ki/ → [seneki] ‘year-REL’ (*senei)

b. /ada-dA-ki/ → [adadaki] ‘island-LOC-REL’ (*adadai)

The application of velar deletion seems to be confined to polysyllabic nouns, and is usually blocked from occurring with monosyllabic nouns. Polysyllabic nouns in the Turkish Electronic Living Lexicon (TELL: Inkelas, Küntay, Orgun, & Sprouse, 2000). corpus have a deletion rate of overall over 90% whereas monosyllables have a deletion rate of only 3% (Becker, Ketrez, & Nevins, 2011; Pycha, Inkelas & Sprouse, 2007) and Turkish speakers extend this trend to nonce words in a wug test (Becker et al., 2011; Zimmer & Abbott, 1978). Whether a particular lexeme alternates, however, is unpredictable. Becker et al. (2011) posit an analysis which relies on lexically-specific faithfulness constraints cloning for each lexical item where the cloned faithfulness constraint (MAX) that blocks deletion is ranked above the markedness constraint *VKV. It should be noted that even in the context in which velar deletion most readily applies, an analysis still necessitates lexically-specific constraints. Thus, Turkish velar deletion appears to be a less-than-canonical MDEE in the same vein as Finnish assibilation discussed in §1.

The current pattern has its roots in a historical process which spirantised then deleted /g/ in intervocalic and post-vocalic position (Halle, 1979; Inkelas & Orgun, 1995; Lees, 1961; Sezer 1986; Ünal-Logacev, Zygis & Fuchs, 2017). /k/, on the other hand, never went through this change (Inkelas & Orgun, 1995).[8] Currently, word-final /g/ (and voiced stops more generally) only occurs in loanwords (Göksel & Kerslake, 2005, e.g. /katalog/ 'catalog'), although Lewis (2000: 10) reports that these are often pronounced with final [k] and not [g]. As with words with final /k/, words with putative final /g/ also show variability in deletion (Inkelas, 2011; e.g. /katalog/ 'catalog' → [kataloa] 'catalog-DAT', but /patalog/ 'pathologist' → [patalogu] 'pathalogist-ACC').

So unlike Korean palatalisation which historically affected a natural class of coronal stops across laryngeal specification, voiced and voiceless velars underwent different historical changes in Turkish. In Modern Turkish, Sezer (1981) points out that /k/-deletion in stems is confined to what was historically a small set of denominal native vowel-initial suffixes (from Sezer, 1981: 375):

[8] I would also like to thank an anonymous reviewer for pointing this out.

(19) Suffixes that trigger deletion:

a. Copula: ‘1.sg.’: /-Im/ ‘1.pl’: /-Iz/

b. Possessive: /-Im/, /-In/, /-I(n)/, /-ImIz/, /-InIz/

c. Accusative: /-I/

d. Dative: /-E/

e. Resemblative: /-ImsI, -Imtrak/

Sezer (1981) also points out that although /k/-deletion does occur with some verbal suffixes, these suffixes share phonological shape with the nominal suffixes which do trigger deletion.

It should already be clear at this point that although often described as together as MDEEs, Korean palatalisation and Turkish velar deletion do not evince the same alternation patterns. In Korean, palatalisation is, to the best of my knowledge, productive and general insofar as it applies to all suffixes which provide the appropriate phonological environment. In Turkish, however, velar deletion is morphologically restricted to certain suffixes. This raises the question of how exactly these languages might differ in terms of their static phonotactic generalisations. In Korean, while there were indeed stem-internal exceptions to the constraint *TI, there was nonetheless a significant enough under-representation of such sequences such that a reliable phonotactic generalisation could be learned. But how strongly, if at all, is the constraint motivating velar deletion represented in the lexicon of Turkish? In the following sections, I report on the results of a corpus investigation and computational learning simulation of Turkish, complementing our previous examination of Korean.

3.2. TELL corpus

First, we examine the lexical statistics in TELL (Inkelas et al., 2000). TELL contains approximately 30,000 lexemes that were compiled from a variety of existing dictionaries as well as transcribed pronunciations from two speakers of a large proportion of these in various verbal and nominal inflected forms. For current purposes, we will be querying the transcribed roots in the database. In total, the resulting corpus contained 16,757 transcribed roots.[9] O/E values were calculated in the same way as in section 4. This time we were interested in comparing the velar stops [k, g] vs. other stops and affricates in the intervocalic context vs. all other contexts. Given the different historical sound changes affecting [k] and [g], these were examined separately here - Table 5 shows this calculation. Recall that an O/E value of 1 indicates that a particular combination of segments co-occurs at essentially an expected rate of co-occurrence given their independent frequency of occurrence; an O/E of 0 indicates that a particular combination does not occur at all. We note that intervocalic [k]s occur close to the expected rate of occurrence (i.e. O/E is close to 1), whereas intervocalic [g]s are somewhat under-attested, which is not surprising given that historically [g]s were lost in this position. This contrasts with Korean where the O/E value of [TI] sequences was much smaller, and in

[9] I have not filtered out roots with duplicates that correspond to different lexemes in the analysis presented here. Calculations using unique roots yield the same qualitative results. The current analysis is presented to allow for the use of lexical category information.

fact closer to 0. In section 3.3, I further examine if a computational learner assigns any weight to a constraint penalizing intervocalic velars in Turkish.

Table 5. Occurrence of [k] and [g] compared to other stops/affricates in V_V vs. other contexts.

	V_V	Other contexts	Expected % of Cs
[k]	900 (1,071) **15.5%** / *23.5%* O/E = 0.84	4,915 (4,744) **84.5%** / *29.0%* O/E = 1.03	*28.0%*
[g]	100 (232) **7.9%** / *2.6%* O/E = 0.43	1,161 (1,029) **92.1%** / *6.9%* O/E = 1.13	*6.1%*
other stops/affricates	2,826 (2,523) **20.6%** / *73.9%* O/E = 1.12	10,871 (11,174) **79.4%** / *64.1%* O/E = 0.97	*65.9%*
Expected % of occurrence in context	**18.4%**	**81.6%**	

3.3. Learning a phonotactic grammar of Turkish

The learning data for the stimulation is the entire TELL corpus of roots. The learner was fed the feature system in Table A2 in the Appendix using both privative and contrastive feature specification. Palatal stops (palatalized velar stops) were included in the segmental inventory, even though they only contrast in certain environments. In the current simulations, long vowels in the corpus are coded as a sequence of two vowels. The features [back], [high] and [round] were used to classify the 8 vowels: [i, y, ɨ, u, e, ø, a, o].

The phonotactic learner was asked to find trigram constraints with a maximum number of constraints set at 180. As with the Korean simulation, the accuracy threshold for constraints was set at 0.3. Unlike in the Korean simulation, we included a Vowel Tier projection, to allow the learner to discover constraints on vowel harmony (e.g. Clements & Sezer, 1982). All other parameters were set at default. As with the Korean simulation, the learner was initialized with the constraint of interest - *[+syllabic][+dorsal][+syllabic] (i.e. *VKV, following Inkelas & Orgun, 1995) - and we were interested in whether the learner assigns this constraint any weight, penalising words with intervocalic velars.

3.4. Results

As with the Korean simulation in §2.5, the top-weighted constraints induced by the learner were inspected to confirm that these conformed to well-known phonotactic constraints in Turkish. The highest weighted constraints that the learner found were primarily related restrictions on consonant clusters (Table 6). This conforms with the fact that Turkish phonotactics tends to only allow certain consonant clusters (van der Hulst & van de Weijer, 1991). The learner also penalises roots which violate rounding harmony, in particular sequences which involve [ɨ]: a form like *[umɨt] gets a penalty score of 5.17 and a form like *[tɨfus] gets a score of 3.60 (forms from Kirchner, 1993; see also Clements & Sezer, 1982). Most importantly, the learner does not assign any weight to the constraint penalizing intervocalic velars (*VKV), and thus words like [sokak] receive a perfect penalty score of 0 and are predicted to be completely phonotactically well-formed.[10] Note, further that although VgV sequences were somewhat underattested

[10] We might worry that the difference between Korean, where the learner did discover *TI, and Turkish, where it did not discover *VKV, is due to the difference in complexity between the two constraints. That is, for Korean, it was only necessary to search the space of bigram constraints, while Turkish requires searching the larger space of trigram

compared to VkV sequences (Table 5), no constraint (e.g. *VgV) was induced that penalised VgV sequence, and test items with intervocalic [g] received a penalty score of 0. Thus, under similar modelling assumptions, the learner does not learn a phonotactic constraint against intervocalic velars in Turkish, but it does in Korean for [TI].

Table 6. Top weighted constraints learned from TELL corpus

No.	Constraint	Weight	Description
1.	*[+wd_bound][+son.,-syll.][-syll.]	5.253	In effect: word-initial CCs have to be obstruent-initial. e.g. *#rt, *#lp
2.	*[-syll.][+son.,-syll.][-syll.]	4.88	In effect: No triconsonantal clusters, e.g. *plk
3.	*[+del_rel.,-voice][+son.,-syll.][+wd_bound]	4.705	In effect: word-final CCs have to be obstruent-final, *pr#, *#pl
4.	*[-del_rel.,-voice,-ant.][+wd_bound]	4.644	*k^j#
5.	*[-syll.][-cons.,+approx.][+wd_bound]	4.51	*Cj#
6.	*[+wd_bound][-ant.][-syll.]	4.201	In effect: palatal or post-alveolar segments cannot be the first C in a word-initial CC. e.g., *ʃt
7.	*[+wd_bound][-syll.][-cons.,+approx.]	4.199	*#Cj
8.	*[-cont.,+voice][-voice]	4.119	In effect: bans sequences of voiced stops/affricates and voiceless segments
9.	*[+wd_bound][+del_rel.,+voice][-syll.]	4.101	In effect: voiced fricatives cannot be the first C in a word-initial CC. e.g., *zr
10.	*[-del_rel.,-voice,-ant.][-back]	4.034	In effect: [k^j] only occurs before back vowels.

3.5. Island of reliability with polysyllabic nouns?

Given the finding that Turkish speakers extend velar deletion to nonce *polysyllabic* words (Zimmer & Abbott, 1978; Becker et al., 2011), there is the possibility that there is an

---

constraints. As a comparison, a trigram model was run with the Korean NAKL corpus, and the same results were obtained.

island of reliability for a phonotactic generalisation (Albright, 2002). That is, there might possibly be a cophonology (Inkelas & Zoll, 2007) or sublexicon (Becker & Gouskova, 2016) of polysyllabic nouns in which there is a strong and reliable phonotactic generalisation (*VKV). To investigate this, I identified polysyllabic nouns in TELL and calculated the same O/E values (Table 7). The same results are obtained as in Table 5 when entire corpus was considered. Given the failure to find a phonotactic constraint in the previous section, it is unlikely that one would be found here either. Thus, it seems that although Turkish speakers do seem to generalize velar deletion to novel polysyllabic nouns (Zimmer & Abbott, 1978), they do not seem to be relying on a general phonotactic generalisation across the entire lexicon or within the sublexicon of polysyllabic nouns.

Table 7. Occurrence of [k] and [g] compared to other stops/affricates in V_V vs. other contexts in Turkish polysyllabic nouns.

	V_V	Other contexts	Expected % of Cs
[k]	644 (765) **18.8%** / *23.0%* O/E = 0.84	2,900 (2,780) **82.8%** / *28.6%* O/E = 1.03	*27.4%*
[g]	77 (157) **10.6%** / *2.8%* O/E = 0.49	651 (571) **89.4%** / *6.4%* O/E = 1.14	*5.6%*
other stops/affricates	2,074 (1873) **23.9%** / *74.2%* O/E = 1.11	6,605 (6,806) **76.1%** / *65.0%* O/E = 0.97	*67.0%*
	21.6%	**78.4%**	

4. Analysis with indexed constraints

Having established the phonotactic generalisations (or lack thereof) available to a computational learner in both Korean and Turkish, in this section, I present an analysis of both the Korean and Turkish cases using indexed constraints (e.g. Becker et al., 2011; Itô & Mester, 1995, 1999, 2001, 2009; Moore-Cantwell & Pater, 2016; Pater, 2007; Smith, 2001). The key point of divergence between the Korean and Turkish is the locus of exceptionality, in particular, whether the alternation is supported by a general markedness (phonotactic) constraint is in the language. In Korean, there is a general markedness constraint (learnable from phonotactic learning), and the exceptional behavior of stems is captured by high-ranked indexed faithfulness. In the Turkish case, however, there is no general markedness constraint, and exceptional triggering of deletion is a result of a high-ranked indexed markedness constraint. Thus, in order to derive the difference in patterns one has to allow utilise indexed markedness (exceptional triggering) and indexed faithfulness constraints (exceptional non-undergoing) (cf. Pater, 2007).

4.1. Korean

Both the corpus studies and phonotactic modeling simulations for Korean in §2 suggest strongly that there is a *general* structure-blind (Martin, 2007, 2011) markedness constraint, *TI, which is active in the Korean grammar. Regardless of whether a morpheme boundary intervenes between T and I, the constraint still penalises the sequence. Yet words with such sequences do exist in the Korean lexicon, e.g. *mati* ‘joint’. An analysis of the Korean palatalisation pattern, therefore, needs to predict that *existing* [TI] words should surface faithfully but it should also penalize *novel* words that do contain such sequences. The analysis presented in this paper builds on the new observation presented in the previous sections. An analysis of Korean palatalisation should:

(i) capture the fact that [TI] sequences are under-represented (=less phonotactically well-formed). Under the analysis here, stem-internal sequences are treated as lexical exceptions given their rarity in the Korean lexicon as a whole, especially within the native and Sino-Korean strata.

(ii) capture the fact that alternations at the morpheme boundary are categorical.

(iii) capture the fact that existing words have fixed outputs containing stem-internal [TI].

To the best of my knowledge, palatalisation across suffix boundaries is both general in that it applies to all suffixes where the phonological conditions are met (20) and extends to loanwords (Jun & Lee, 2007).

(20) Palatalisation occurs in both inflectional and derivational suffixes (data from Cho, 2009)

a.	/hɛ tot-i/ sun rise-NML	→	[hɛ toci]	‘sun-rise’
b.	/kut-i/ be.firm-ADV	→	[kuci]	‘firmly’
c.	/put^{h}-i/ adhere-CAUS	→	[puc^{h}i]	‘to affix’
d.	/pat^{h}-ita/ field-COP	→	[pac^{h}ita]	‘to be the field’

While the computational learner penalizes tautomorphemic [TI] sequences by assigning forms with such sequences a non-zero penalty score, the penalty assigned is weaker than what is essentially a categorical constraint (i.e. completely ill-formed), such as *#ŋ, indicating that this is a gradient restriction (e.g. Coetzee & Pater 2008). For current purposes, I set aside this issue of gradience to highlight the most salient difference between the Korean and Turkish cases discussed here. I return to this issue below. While the analysis here is presented using Harmonic Grammar (Legendre, Miyata & Smolensky, 1990), the main thrust of the analysis does not hinge on this analytical choice.

Given that the computational learner penalized [TI] sequences, I propose that the grammar contains a general structure-blind markedness constraint which penalises [TI] sequences both within a morpheme and across a morpheme boundary (21).

(21) *TI:

Assign one violation mark to every sequence of [t, t^h, t*][i, j] in the output.

This markedness constraint is in competition with a faithfulness constraint that prevent the alternation. Here, I just assume a general FAITH constraint that encompasses all possible repairs. (22) shows that *TI has a higher weight than FAITH to account for the alternation at the morpheme boundary.

(22)

/mat-i/ 'eldest-NOM'	*TI *w* = 2	FAITH *w* = 1	$\mathcal{H}$
a. [mat-i]	-1		-2
☞ b. [mac-i]		-1	-1

In order to ensure that existing words like *mati* 'joint' surface faithfully, I propose an additional, lexically-specific indexed faithfulness constraint along the lines of Burzio

(2000; see also the discussion in Pater, 2007; and Itô & Mester's (1995) NEIGHBORHOOD schema) prevents palatalisation in such cases. This constraint is indexed to existing stems that contain [TI] sequences, e.g. *mati* 'joint', along the lines of Moore-Cantwell and Pater (2016), and prevents alternations in existing words as shown in (23).

(23)

/mati/ 'joint'	FAITH$_{TI,mati}$ $w = 3$	*TI $w = 2$	FAITH $w = 1$	$\mathcal{H}$
☞ a. [mati]		-1		-2
b. [maci]	-1		-1	-4

Under the proposed analysis, non-application of Korean palatalization in specific stems is an instance of exceptional non-undergoers due to high-ranked indexed faithfulness constraints.

The current analysis predicts that novel words with [TI] should palatalise, although a reviewer points out, as does Cho (2009), that loanwords with [TI] sequences in Korean do not undergo palatalisation. In accordance with this, our corpus investigation above in §2.3 revealed that nearly two-thirds of [TI] words are in the loanword stratum of the lexicon. The analysis here is in principle compatible with these patterns if we assume a faithfulness constraint indexed to loanwords that outranks the general markedness constraint (e.g. Itô & Mester, 1999, 2009). An alternative possibility, also raised by a reviewer, is that Korean speakers might simply not be aware of the rarity of /TI/ in the lexicon. If this is true, this would raise a further interesting question of why this should be given that palatalisation is a phonetically natural pattern (cf. Becker et al, 2011), and we know that speakers can track statistical regularities in phonotactic patterns, and based on these, show gradient well-formedness intuitions regarding novel words in their language (e.g., Frisch & Zawaydeh, 2001; Hay et al., 2003; Hayes & White, 2013). Future experimental examination of native speaker knowledge is therefore needed to examine these possibilities.[11]

The analysis above also does not address the fact that while [TI] sequences are systematically penalized by a general markedness constraint, it is not penalized to the same degree as what we know to be categorical phonotactic constraints (e.g. no [ŋ] at the start of a word). This presents a case of gradient exceptionality, and how this is learnt in conjunction with categorical alternations is unclear. Thus, in addition to native speaker judgements, we would also need a model of how constraint weights are learnt in the face of gradient exceptionality in the lexicon, and how this impacts the productivity of a phonological alternation. This is a subject of ongoing research (e.g., Moore-Cantwell & Pater, 2016; Zymet, 2018), and is beyond the scope of the current paper.

The core claim of the analysis here, based on the results of the phonotactic learning simulation, is that the putative MDEE pattern in Korean palatalisation is a case exceptional non-undergoers, in violation of a more general markedness constraint. The crucial distinction is not simply between a derived and non-derived (stem-internal) environment, since the indexed faithfulness constraints only apply to specific lexical items and not the stem-internal domain as a whole. In this way, the analysis presented here departs from previous proposals (e.g. constraint conjunction, Łubowicz, 2002; sequential faithfulness, Bradley, 2007; Burzio, 2000; T. Cho, 2001; Itô & Mester, 1998;

[11] It is a puzzle as to why Korean allowed [TI] sequences to be borrowed in faithfully in the first place. Loaning occurred after the counterfeeding diachronic sound change was complete (Cho, 2009), producing novel [TI] sequences. Yet as we saw from the corpus results, TI was strongly under-represented in the native and Sino-Korean lexicons. It is possible here that other considerations regarding loanword adaptation are at play which prefer such forms (e.g. orthographic effects: Daland, Oh, & Kim, 2015).

Wolf, 2008; underspecification, Kiparsky, 1973; strength scales, Inkelas, 2015), and predicts that nonce words with [TI] are predicted to be less acceptable to native Korean speakers when compared to words with other sequences such as [ta] or [ci]. Previous accounts predict that such nonce words should be perfectly acceptable. The claim here, however, is that the stem-internal TI sequences are exceptions to the more general markedness constraint that disprefers such sequences. Further confirmation of this prediction awaits future work testing native speaker knowledge.

4.2. Turkish

Having proposed an analysis using indexed faithfulness constraints of the Korean MDEE pattern in the previous section, I now turn to an analysis of Turkish velar deletion. As described above and elsewhere, the alternation occurs for the most part only with nominal vowel-initial suffixes (Inkelas, 2011, Inkelas & Orgun, 1995; Sezer 1981); suffix-initial /k/s do not delete. Inkelas & Orgun (1995) accounted for this by having the constraint against intervocalic velars apply at only a specific morphological level in the derivation. More recently, Becker et al. (2011) examined the alternation rates for velar-final noun stems in TELL. Their analysis however was restricted to one morphological environment: the possessive (see also Zimmer & Abbott, 1978). In their wug study, Turkish participants applied the deletion rule in the possessive context to polysyllabic nonce words and not monosyllabic nonsense words, in accords with the trends in the lexicon. The TELL corpus, however, also includes tokens in two other morphological contexts that should trigger deletion: the accusative and predicative. Two further observations can be made from an examination of deletion rates in these contexts. Firstly, although deletion is by far the most common pattern here, each suffix elicited seems to have its own rate of deletion, in what Zymet has recently terms ‘lexical propensities’ (2018): predicative (78%), accusative (91%) and possessive (93%). Secondly, a given root does not always behave the same way across morphological contexts. For example, a root that Becker et al. (2011) classify as an alternator /ajak/ ‘foot’, does indeed show the alternation in the possessive form [aja-ɯm] and accusative [aja-ɯ], but not in the predicative form [ajak-ɯm].

Here, I assume Becker et al.’s (2011) analysis as a starting point. Given the observations above, and the fact that Turkish does not have a general phonotactic constraint available across the entire lexicon, I propose a small modification to their analysis to account for the fact that the constraint motivating deletion is only associated with a small set of nominal suffixes, enumerated in Sezer (1981; see also Inkelas 2011). Becker et al.’s (2011) analysis uses faithfulness constraints that are lexically-indexed to the roots that do not show the alternation, resulting in a constraint ranking like $\textsc{Max}_i$ >> *VKV >> MAX. The suggestion here is a modest one and has to do with ensuring that the *VKV constraint (24) is indexed to specific individual triggering suffixes (Pater, 2007) where the right edge of this configuration’s domain must overlap with an exponent of the relevant morphemes that trigger the alternation (Jarosz, 2018). Given that each suffix shows its own rate of triggering deletion, I suggest there that there is a cloned markedness constraint indexed to each one.

(24) $*VKV]_{possessive}$

Assign one violation mark for every sequence of VKV where the right edge of the configuration's domain overlaps with some exponent of the possessive morpheme.

The constraint is only violated if the right edge of the configuration's domain overlaps with some exponent of the possessive morpheme; otherwise, it does not apply. Thus for a root like /ajak/ in the possessive alternates under the constraint ranking $*VKV]_{poss}$ >> $\textsc{Max}_{ajak}$, but not in the predicative under the ranking $\textsc{Max}_{ajak}$ >> $*VKV]_{pred}$. In both cases, the structure-blind markedness constraint *VKV is low-ranked or receives no weight, unlike in Korean.

The approach here captures the idea that that the phonotactic constraint is not a general one within the language. It also captures the observation that the alternation is triggered only by specific suffixes, which seemingly have their own rate of application, and the roots themselves can have various rates of blocking. Whether the suffix-specific rates of trigger deletion hold up once we examine all potential contexts and across a larger corpus of data remains to be determined. Regardless, in Turkish, in contrast to Korean, the markedness constraint has to be indexed to a subset of suffixes only.

Clearly, given the results from both Zimmer & Abbott (1978) and Becker et al. (2011), velar deletion in Turkish is productive in the relevant morphological contexts, since the markedness constraint is indexed to specific triggering suffixes. Turkish speakers, in fact, reproduced the frequency of alternations based on the phonological factors that Becker et al. (2011) examined. If the learning of phonotactics both precedes and facilitates the learning of alternations (Hayes, 2004; Tesar & Prince, 2007; Pater &

Tessier, 2005; Hayes & Wilson, 2008), then we expect that Turkish velar deletion would be learnt later than an alternation that is phonotactically supported, like Korean palatalisation, only once more sophisticated knowledge of morphology and paradigms is in place.

5. General Discussion

The present paper set out with the aim of examining in closer detail the quality of the phonotactic generalisations available in languages with derived-environment effect patterns. In particular, I examined whether stem-internal sequences are truly grammatically well-formed in derived-environment effect patterns by comparing stem-internal phonotactic generalisations in Korean palatalisation and Turkish velar deletion, two well-known MDEE cases. While there is a robust, albeit gradient, phonotactic constraint against [TI] that is able to motivate alternations in Korean, no such constraint against intervocalic velars is readily available in Turkish. Both languages also differ on the extent to which the alternation is morphologically-general: in Korean any suffix that meets the phonological requirements triggers the alternation, while in Turkish only a subset of suffixes trigger the alternation (Sezer, 1981; Inkelas, 2011). Here I discuss the implications of these results for MDEEs, and what this means for theoretical analyses of these patterns and MDEEs as a unified phenomenon more generally.

5.1. MDEEs with indexed constraints

The analyses presented above utilises indexed constraints, specifically both indexed markedness and indexed faithfulness constraints. The use of indexed constraints to account for MDEEs is certainly not new (e.g. Burzio, 2000; Pater, 2007; Jurgec & Bjorkman, 2018). Here I am suggesting that both indexed faithfulness and indexed markedness constraints are needed to account for the different patterns. The main difference between the analysis of the patterns in Korean and Turkish is the extent to which a general structure-blind phonotactic markedness constraint is active in the grammar. In the Turkish case, given that the markedness constraint is indexed to specific suffixes (and not just suffixes in general), this is an instance of suffix-specific (i.e. exceptional) triggering by a markedness constraint that is not active elsewhere. This analysis is further supported by the learning simulations which show that a general phonotactic constraint against intervocalic velars is not available. In contrast, the computational learner does assign a non-trivial weight to a general phonotactic constraint against [TI] sequences in Korean. To capture this, I proposed that a structure-blind (non-indexed) markedness constraint is responsible for penalising [TI] sequences in Korean, but palatalisation is blocked through a lexically-specific faithfulness constraint to ensure that existing words with [TI] stem-internally surface faithfully.

Unlike traditional accounts of MDEEs, I have not attempted here to provide a unified analysis of both patterns, reflecting the fact that while superficially similar, the actual phonological patterns in each case are not identical, a point I take up in the next section.

5.2. MDEEs as a unified phenomenon?

A major claim of the current study is that despite surface similarities, patterns previously described together as examples of MDEE are by no means a unified phenomenon. On the surface, both Korean palatalisation and Turkish velar deletion (and in fact, MDEEs in general) share similarities in that the phonological process is purported to only apply when the environment is achieved by virtue of the concatenation of two morphemes, but not within morphemes. This surface similarity belies major differences once we start looking more closely at the quantitative patterns in the lexicon and with alternations. On the one hand, Korean palatalisation is categorical across a morpheme boundary, and not confined to specific suffixes, while stem-internal forms are rare enough that they receive a well-formedness penalty. On the other hand, Turkish velar deletion, while productive, is confined to specific suffixes, and unlike Korean, intervocalic velars within stems are entirely phonotactically well-formed as I showed above.

In fact, while both patterns are traditionally considered uncontroversial textbook examples of MDEEs, they do not hold up to scrutiny, especially when one takes into consideration the assumptions laid out in (5) and (11) repeated here as (25) and (26).

(25) DERIVED-ENVIRONMENT CONDITION:

Morphological derivedness is a necessary and *sufficient* condition for a process to occur (variously stated as the Strict Cycle Condition or the Revised Alternation condition; Kiparsky 1973, Kiparsky 1982).

(26) PHONOTACTIC WELL-FORMEDNESS:

Morpheme-internal instances of sequences repaired at the morphological boundary are phonotactically well-formed.

That Turkish fails on (25) has previously been pointed out by Inkelas (2011; cf. Sezer, 1981). Further as discussed in §1, Finnish assibilation (Kiparsky, 1973, 1993; Anttila 2006), perhaps the most-cited example of MDEEs likewise fails on the derived-environment condition. Observations such as these led Inkelas (2011) to argue more generally that the derived-environment condition is not a sufficient, or potentially useful, diagnostic for putative MDEE patterns. Rather, she conjectures that many cases previously described as MDEEs are in fact just cases of morphologically-conditioned phonology.

Korean, on the other hand, does not satisfy (26): stem-internal sequences are not in fact completely phonotactically well-formed[12]. Thus, at least some putative cases of MDEE are instances of lexical exceptionality. Our corpus and phonotactic modelling simulation quantitatively support Cho's (2009) initial observation of Korean. In Korean, stem-internal [TI] sequences are exceedingly rare and come mostly from loanwords. Furthermore, Cho (2009) points out that some other cases of MDEEs often involve loanwords (e.g. Finnish Vowel Coalescence (Anttila, 2009), Polish First Velar

[12] Korean arguably also violates (25) if we take into account compounding, where underlying /ti/ sequences can also occur across compound (or prefix) boundaries (Oh, 1995). But there is a conspiracy here where n-insertion can variably occur to fix such sequences. It is an open question then as to the extent to which *n*-insertion applies to prevent underlying /ti/ from surfacing.

Palatalisation (Łubowicz, 2002)). A closer inspection of the lexicon of at least some well-known MDEE cases might well reveal a similar picture as in Korean where we observe a phonotactic dispreference within-stems for sequences repaired across morpheme boundaries.

Taken together, the strongest claim that can be made is that MDEEs as a formal category should be dispensed with (for a similar conclusion, see Inkelas, 2011). Neither of the cases examined in detail here fulfil the traditional criteria for MDEEs, and they fail in different ways, requiring different theoretical treatment. One presents a case of exceptional triggering (high-ranked indexed markedness), and another exceptional non-undergoers (high-ranked indexed faithfulness). The analyses presented above do, however, do not rule out the 'true' MDEE cases of categorical alternation at the morpheme boundary but where stem-internal sequences are entirely well-formed. In principle, a markedness constraint indexed to affixes in general, as suggested by Pater (2007), would be able to account for these traditional patterns. This would not certainly require any theoretical machinery specific to MDEEs, but would fall out from existing theoretical tools of constraint indexation (for a recent extension of indexed constraints, see Jurgec & Bjorkman, 2018; cf. Rasin, 2016). This might be desirable since, as a reviewer points out, it is entirely possible that such 'classical' cases exist but have yet to be uncovered. But, these 'classical' MDEEs would just be a specific case of morpheme-specific phonology, alongside the cases discussed above. Ultimately, however, if upon further examination, more cases end up patterning like the ones observed here, it might be necessary to modify the analysis to rule out 'classical' cases completely.

If MDEEs as a notion is to be retained, it certainly cannot be viewed as a unified phenomenon accounted for by a single analysis, suggesting the possibility of a 'typology' of MDEE patterns which share in common the fact that a particular phonological process applies in some morphological contexts but not others. Such an expanded version of MDEEs is referred to in a recent paper by Jurgec and Bjorkman (2018: 577) who include phonological processes applying to "only a subset of morphemes (e.g. to loanwords, or to affixes but not roots)" as instances of MDEEs. This is already an expansion of the traditional notion of MDEEs. A more in-depth understanding of patterns previously described as MDEEs is required to ultimately provide a better understanding of the extent to which these patterns differ from each other, and whether all putative MDEE patterns reduce to either exceptional triggers vs. exceptional undergoers as is suggested here? If so, this would raise the question of whether it is even useful to consider MDEEs as a special category of morphophonological patterns, distinct from morphologically-conditioned phonology or lexical exceptionality more generally.

I leave the ultimate answer to these issues open for future examination. At the very least, beyond identifying whether a pattern occurs in some morphological domain, but not the other (e.g. within stems), we should also examine whether that particular pattern is indeed categorical at all, and if so to what extent. The examination of these patterns in more detail, including from an acquisition and diachronic perspective, will also prove a fertile ground for investigating the extent to which stem phonotactics and alternations can pull apart .

To conclude, in this paper, I showed that contrary to existing assumptions, MDEE patterns do not form a uniform phenomenon. I showed that the difference, in at least the cases examined here, lies in the locus of exceptionality. In Korean, there is an active alternation with a gradient static phonotactic constraint, whereas in Turkish, there is a morphologically-conditioned alternation with no corresponding static phonotactic constraint. Ultimately, these results call into question a unified notion of MDEEs, and further, whether MDEEs are even a useful notion to apply to the analysis of phonological patterns.

Appendix

Table A1. Feature system used for Korean modelling simulation (based on Cho, 2012)

	syllabic	consonantal	sonorant	continuant	aspirated	tense-cg	nasal	lateral	spread_glot	labial	coronal	anterior	strident	dorsal	high	low	back	round
p	-	+	-	-	-	-				+								
p^h	-	+	-	-	+	-				+								
p*	-	+	-	-	-	+				+								
t	-	+	-	-	-	-					+	+	-					
t^h	-	+	-	-	+	-					+	+	-					
t*	-	+	-	-	-	+					+	+	-					
c	-	+	-	-	-	-					+	-	+					
c^h	-	+	-	-	+	-					+	-	+					
c*	-	+	-	-	-	+					+	-	+					
k	-	+	-	-	-	-								+				
k^h	-	+	-	-	+	-								+				
k*	-	+	-	-	-	+								+				
s	-	+	-	+	-	-					+	+	+					
s*	-	+	-	+	-	+					+	+	+					
h	-	+	-	+					+									
m	-	+	+				+			+								
n	-	+	+				+				+	+						
ŋ	-	+	+				+							+				
l	-	+	+					+			+	+						
j	-	-	+												+	-	-	-
w	-	-	+							+					+	-	+	+
i	+	-	+												+	-	-	-
e	+	-	+												-	-	-	-
ɨ	+	-	+												+	-	+	-
ʌ	+	-	+												-	-	+	-
a	+	-	+												-	+	+	-
o	+	-	+												-	-	+	+
u	+	-	+												+	-	+	+

Table A2. Feature system used for Turkish modelling simulation

	syllabic	consonantal	sonorant	continuant	delayed_release	approximant	tap	nasal	voice	spread_glot	labial	labiodental	coronal	anterior	distributed	strident	lateral	dorsal	high	back	round
p	-	+	-	-	-				-		+										
b	-	+	-	-	-				+		+										
t	-	+	-	-	-				-				+	+	-	-					
d	-	+	-	-	-				+				+	+	-	-					
c	-	+	-	-	-				-				+	-	+	-					
ɟ	-	+	-	-	-				+				+	-	+	-					
k	-	+	-	-	-				-									+			
g	-	+	-	-	-				+									+			
tʃ	-	+	-	-	+				-				+	-	+	+					
dʒ	-	+	-	-	+				+				+	-	+	+					
f	-	+	-	+	+				-		+	+									
v	-	+	-	+	+				+		+	+									
s	-	+	-	+	+				-				+	+	-	+					
z	-	+	-	+	+				+				+	+	-	+					
ʃ	-	+	-	+	+				-				+	-	+	+					
ʒ	-	+	-	+	+				+				+	-	+	+					
h	-	+	-	+	+				-	+											
m	-	+	+					+			+										
n	-	+	+					+					+	+	-	-					
ɾ	-	+	+			+	+						+	+	-	-					
l	-	+	+			+							+	+	-	-	+				
j	-	-	+			+													+	-	-
i	+	-	+																+	-	-
y	+	-	+																+	-	+
ɯ	+	-	+																+	+	-
u	+	-	+																+	+	+
e	+	-	+																-	-	-
ø	+	-	+																-	-	+
o	+	-	+																-	+	-
a	+	-	+																-	+	+

References

Albright, A. (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language*, *78*(4), 684–709.

Anttila, A. (2006). Variation and opacity. *Natural Language & Linguistic Theory*, *24*, 893–944.

Anttila, A. (2009). Derived Environment Effects in Colloquial Helsinki Finnish. In K. Hanson & S. Inkelas (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 433–460). Cambridge: MIT Press.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2) [CD-ROM]*. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium [Distributor].

Bailey, T. M., & Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory and Language*, *44*(4), 568–591.

Becker, M., & Gouskova, M. (2016). Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry*, *47*(3), 391-425.

Becker, M., Ketrez, N., & Nevins, A. (2011). The Surfeit of the Stimulus: Analytic Biases Filter Lexical Statistics in Turkish Laryngeal Alternations. *Language*, *87*(1), 84–125.

Bradley, T. J. (2007). Morphological derived-environment effects in gestural coordination: A case study of Norwegian clusters. *Lingua*, *117*(6), 950–985.

Burzio, L. (2000). Cycles, non-derived environment blocking, and correspondence. In:

J. Dekkers, F. van der Leeuw and J. van de Weijer (eds.), *Optimality Theory: Phonology, syntax and acquisition* (pp. 47-87). Oxford: Oxford University Press.

Burzio, L. (2011). Derived Environment Effects. In M. Oostendorp, C. J. Ewen, E. Hume and K. Rice (eds.), *The Blackwell Companion to Phonology*. doi:10.1002/9781444335262.wbctp0088

Cho, T. (2001). Effects of Morpheme Boundaries on Intergestural Timing: Evidence from Korean. *Phonetica*, *58*(3), 129–162.

Cho, N. (2003). *Hangwukeo Hakseupyong Eohwi Seonceong Kjeolkwa Pogoseo.* The National Academy of the Korean Language. [http://www.korean.go.kr]

Cho, Y. Y. (2009). A historical perspective on nonderived environment blocking: The case of Korean palatalisation. In K. Hanson & S. Inkelas (Eds.), *The Nature of the Word: Studies in Honor of Paul Kiparsky* (pp. 461–486). Cambridge: MIT Press.

Cho, H. (2012). Statistical learning of Korean phonotactics. *Studies in Phonetics, Phonology and Morphology, 18*(2), 339-370.

Chong, J.A. (2017). On the relation between phonotactic and alternation learning (Ph.D. Dissertation). University of California, Los Angeles, Los Angeles, CA.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York, NY: Harper and Row.

Chung, S. (1983). Transderivational Relationships in Chamorro Phonology. *Language*, *59*(1), 35.

Clements, G.N. & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. In H. van der Hulst and Norval Smith (eds.) *The structure of phonological representations, II* (pp.213-254). Dordrecht: Foris.

Coetzee, A. W., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, *26*(2), 289–337.

Coleman, J., & Pierrehumbert, J. B. (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology: Third meeting of the ACL special interest group in computational phonology* (pp. 49–56). Association for Computational Linguistics. Retrieved from http://arxiv.org/abs/cmp-lg/9707017

Daland, R., Oh, M., & Kim, S. (2015). When in doubt, read the instructions: Orthographic effects in loanword adaptation. *Lingua*, *159*, 70–92.

Eychenne, J., & Jang, T.-Y. (2015). On the merger of Korean mid front vowels: Phonetic and phonological evidence. *Journal of the Korean Society of Speech Sciences*, *7*(2), 119–129.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords. *Journal of Memory and Language*, *42*(4), 481–496.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, *22*(1), 179–228.

Frisch, S. A., & Zawaydeh, B. A. (2001). The psychological reality of OCP-Place in Arabic. *Language*, *77*(1), 91–106.

Göksel, A., & Kerslake, C. (2005). *Turkish: a comprehensive grammar*. London: Routledge.

Halle, M. (1979). Formal vs. functional considerations in phonology. In B. Brogyanyi (ed.), *Studies in diachronic, synchronic and typological linguistics* (pp. 325-341). Amsterdam: John Benjamis.

Hammond, M. (1992). Deriving the Strict Cycle Condition. In *Papers from the 28th Regional Meeting of the Chicago Linguistic Society* (pp. 126–140). Chicago: Chicago Linguistic Society.

Hay, J., Pierrehumbert, J. B., & Beckman, M. (2003). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 58–74). Cambridge: Cambridge University Press.

Hayes, B. (2000). Gradient well-formedness in optimality theory. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (Eds.), *Optimality Theory: Phonology, Syntax, and Acquisition* (pp. 88–120). Oxford: Oxford University Press.

Hayes, Bruce (2004). Phonological acquisition in Optimality Theory: the early stages. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition* (pp. 158–203). Cambridge: Cambridge University Press.

Hayes, B., Siptár, P., Zuraw, K., & Londe, Z. (2009). Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language*, *85*(4), 822–863.

Hayes, B., & White, J. (2013). Phonological Naturalness and Phonotactic Learning. *Linguistic Inquiry*, *44*(1), 45–75.

Hayes, B., & Wilson, C. (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, *39*(3), 379–440.

Inkelas, S. (2000). Phonotactic blocking through structural immunity. In B. Stiebels & D. Wunderlich (Eds.), *Lexicon in focus* (pp. 7–40). Berlin: Akademie Verlag.

Inkelas, S. (2011). Another look at velar deletion in Turkish, with special attention to the derived environment condition. In E. E. Taylan & B. Rona (Eds.), *Puzzles of Language: Essays in honour of Karl Zimmer* (Turcologica 86). Wiesbaden: Harrassowitz.

Inkelas, S. (2015). Confidence scales: a new approach to derived environment effects. In Y. E. Hsiao & L.-H. Wee (Eds.), *Capturing phonological Shades* (pp.45-75). Cambridge: Cambridge Scholar Press.

Inkelas, Sharon, Aylin Küntay, Ronald Sprouse, and Cemil O. Orgun. (2001). TELL: Turkish Electronic Living Lexicon. *Turkic Languages*, 4(2), 253-75.

Inkelas, S., & Orgun, C. O. (1995). Level Ordering and Economy in the Lexical Phonology of Turkish. *Language*, *71*(4), 763–793.

Inkelas, S., Orgun, C. O., & Zoll, C. (1997). Implications of lexical exceptions for the nature of the grammar. In I. Roca (Ed.), *Derivations and constraints in phonology* (pp. 393-418). Oxford: Clarendon Press.

Inkelas, S., & Zoll, C. (2007). Is Grammar Dependence Real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics*, *45*, 133–171.

Itô, J., & Mester, A. (1995). The Core-periphery Structure in the Lexicon and Constraints on Re-ranking. In J. Beckman, L.W. Dickey & S. Urbanczyk (eds.), *Papers in Optimality Theory* (pp.181-210). GLSA, University of Massachusetts, Amherst.

Itô, J., & Mester, A. (1998). *Markedness and word structure: OCP effects in Japanese*. Retrieved from http://roa.rutgers.edu/files/255-0498/roa-255-ito-2.pdf

Itô, J., & Mester, A. (1999). The structure of the phonological lexicon. In N. Tsujimura (ed.), *The Handbook of Japanese Linguistics* (pp. 62-100). Oxford: Blackwell.

Itô, J., & Mester, A. (2001). Covert generalizations in Optimality Theory: the role of stratal faithfulness constraints. *Studies in Phonetics, Phonology, and Morphology*, *7*, 273-299.

Itô, J., & Mester, A. (2009). Lexical classes in phonology. In S. Miyagawa & Saito, M. (eds.), *Handbook of Japanese Linguistics* (pp.84-106). Oxford: Oxford University Press.

Itô, J., Mester, A., & Padgett, J. (1995). Licensing and underspecification in Optimality Theory. *Linguistic Inquiry*, *26*(4), 571–631.

Iverson, G. K., & Wheeler, D. W. (1988). Blocking and the Elsewhere condition. In M. Hammond & M. Noonan (Eds.), *Theoretical morphology: Approaches in modern linguistics* (pp. 325–338). San Diego: Academic Press.

Jarosz, G. (2018). Indexed morphemes and locality conditions on Polish yers deletion. Talk presented at 26 Manchester Phonology Meeting.

Jun, J., & Lee, J. (2007). Multiple stem-final variants in Korean native nouns and loanwords. *Journal of the Linguistic Society of Korea [Eoneohag]*, *47*, 159–187.

Jurgec, P. & Bjorkman, B. (2018). Indexation to stems and words. *Phonology, 35*(4), 566-615.

Kager, R., & Shatzman, K. (2007). Phonological constraints in speech processing. *Linguistics in the Netherlands*, *24*, 99–111.

Karlsson, F. (1983). *Suomen kielen ¨a¨anne- ja muotorakenne [The Phonological and Morphological Structure of Finnish]*. Helsinki: Werner Söderström Osakeyhtiö.

Kenstowicz, M., & Kisseberth, C. (1977). *Topics in phonological theory*. New York, NY: Academic Press.

Kenstowicz, M., & Kisseberth, C. (1979). *Generative Phonology: Description and theory*. New York, NY: Academic Press.

Kim, B., Lee, G., & Lee, J.-H. (2002). Morpheme-based grapheme to phoneme conversionusing phonetic patterns and morphophonemic connectivity information. In *ACM Trans. Asian Lang. Inf. Process* (Vol. 1, pp. 65–82).

Kirchner, R. (1993). *Turkish vowel harmony and disharmony: An Optimality Theoretic account*. Talk presented at Rutgers Optimality Workshop I. Available from Rutgers Optimality Archive, # 4.

Kiparsky, P. (1973). Phonological representations. In O. Fujimura (Ed.), *Three dimensions of linguistic theory* (pp. 1–135). Tokyo: TEC.

Kiparsky, P. (1982). *Explanation in Phonology*. Dordrecht: Foris.

Kiparsky, P. (1993). Blocking in nonderived environments. In E. Kaisse & S. Hargus (Eds.), *Phonetics and Phonology 4: Studies in lexical phonology* (pp. 277–313). San Diego: Academic Press.

Kiparsky, P. (2003). Finnish noun inflection. In D. Nelson & S. Manninen (Eds.), *Generative Approaches to Finnic Linguistics* (pp. 109–161). Stanford: CSLI Publications.

Lantz, B. (2013). The large sample size fallacy. *Scandinavian Journal of Caring Sciences,* 27, 487-492.

Lee, K.-M., & Ramsey, S. R. (2011). *A history of the Korean Language*. Cambridge: Cambridge University Press.

Lees, R.B. (1961). *The phonology of modern standard Turkish.* Indiana University Publications: Bloomington.

Legendre, G., Miyata, Y. & Smolensky, P. (1990). Harmonic Grammar – a Formal Multi-Level Connectionist Theory of Linguistic Wellformedness: An Application. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 884–891). Cambridge, MA: Lawrence Erlbaum.

Lewis, G. (1967). *Turkish grammar*. Oxford: Oxford University Press.

Łubowicz, A. (2002). Derived environment effects in Optimality Theory. *Lingua*, *112*(4), 243–280.

Martin, A. (2007). *The evolving lexicon* (Ph.D. Dissertation). University of California, Los Angeles, Los Angeles, CA.

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

McCarthy, J. J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, J. J. (2003). Comparative markedness. *Theoretical Linguistics*, *29*, 1–51.

Meeusen, A. E. (1959). *Essai de grammaire rundi*. Tervuren: Musée Royal du Congo Belge.

Moore-Cantwell, C., & Pater, J. (2016). Gradient exceptionality in maximum entropy grammar with lexically specific constraints. *Catalan Journal of Linguistics*, *15,* 53-66.

NAKL. (2003). Hyeondae gug-eo sayong bindo josa gyeolgwa pail [Frequently used words in modern language]. National Academy of Korean Language. Retrieved from http://www.korean.go.kr/

Oh, M. (1995). A prosodic analysis of nonderived-environment blocking. *Journal of East Asian Linguistics*, *4*(4), 261–279.

Paster, M. (2013). Rethinking the "duplication problem." *Lingua*, *126*, 78–91.

Pater, J. (2007). The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In L. Bateman, M. O'Keefe, E. Reilly & A. Werle (eds.), *University of Massachusetts Occasional Papers 32: Papers in Optimality Theory III*, 259-296. Amherst: GLSA.

Pater, J., & Tessier, A.-M. (2005). Phonotactics and alternations: Testing the connection with artificial language learning. In K. Flack & S. Kawahara (Eds.), *University of Massachusetts Occasional Papers in Linguisitcs* [*UMOP*] (Vol. 31, pp. 1–16). Amherst: GLSA.

Pierrehumbert, J. B. (1994). Syllable structure and word structure: A study of triconsonantal clusters in English. In P. A. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology Il* (pp. 168–188). Cambridge: Cambridge University Press.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2-3), 115–154.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Cambridge, MA: Blackwell Publishers.

Pycha, A. (2008). *Morphological sources of phonological length.* (Ph.D. Dissertation). University of California, Berkeley.

Pycha, A., Inkelas, S. & Sprouse, R. (2007). Morphophonemics and the lexicon: a case study from Turkish. In M. J. Sole, P. Beddor, and M. Ohala (eds.), *Experimental Approaches to Phonology* (pp. 369-385). Oxford: Oxford University Press.

Rasin, E. (2016). *Morpheme structure constraints and blocking in nonderived environments.* Ms, MIT.

Richtsmeier, P. T. (2011). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology*, *2*(1), 157–183.

Rodegem, F. M. (1970). *Dictionnaire Rundi-français*. Tervuren: Musée Royal de l'Afrique Central.

Rubach, J. (1984). *Cyclic and lexical phonology*. Dordrecht: Foris.

Ryu, J. (2012). The L1 Acquisition of the Imperfective Aspect Markers in Korean: a Comparison with Japanese. *International Journal of Asian Language Processing*, *22*, 1–13.

Sezer, E. (1981). The k/ø alternation in Turkish. In G. N. Clements (Ed.), *Harvard Studies in Phonology* (pp. 354–382). Bloomington: IULC.

Sezer, E. (1986). An autosegmental analysis of compensatory lengthening in Turkish. In L. Wetzels & E. Sezer (eds.), *Studies in compensatory lengthening* (pp. 227-250). Dordrecht: Foris.

Shin, J., Kiaer, J., & Cha, J. (2013). *The sounds of Korean*. Cambridge: Cambridge University Press.

Smith, J.L. (2001). Lexical category and phonological contrast. In R. Kirchner, J. Pater & W. Wikely (eds.), *Petl 6: Proceedings of the workshop on the lexicon in phonetics and phonology*, 61–72. University of Alberta. Available from Rutgers Optimality Archive, # 728.

Sohn, H.-M. (1999). *The Korean language*. Cambridge: Cambridge University Press.

Tesar, B., & Prince, A. (2007). Using phonotactics to learn phonological alternations. In *Proceedings of the Thirty-Ninth Conference of the Chicago Linguistics Society, vol. II: The panels* (pp. 209–237).

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. In M. B. Broe & J. PIerrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 269–282). Cambridge: Cambridge University Press.

Ünal-Logacev, Ö, Zygis, M. & Fuchs, S. (2017). Phonetics and phonology of soft 'g' in Turkish. *Journal of the International Phonetic Association,* 1-24.

van der Hulst, H.G. & van der Weijer, J. (1991). Topics in Turkish phonology. In H. Boeschoten & L. Verhoeven (eds.), *Structure and use of Turkish* (pp.11-59). Leiden: Brill.

Wilson, Colin, and Marieke Obdeyn. 2009. *Simplifying subsidiary theory: Statistical evidence from Arabic, Muna, Shona, and Wargamay*. Ms., Johns Hopkins University.

Wolf, M. (2008). *Optimal Interleaving: Serial Phonology-Morphology Interaction in a Constraint-Based Model* (Ph.D. Dissertation). University of Massachusetts, Amherst.

Zimmer, K. E., & Abbott, B. (1978). The k/ø alternation in Turkish: Some experimental evidence for its productivity. *Journal of Psycholinguistic Research*, *7*(1), 35–46.

Zimmer, K. E., & Orgun, C. O. (1999). Turkish. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp. 154–158). Cambridge: Cambridge University Press.

Zuraw, K. (2000). *Patterned exceptions in phonology* (Ph.D. Dissertation). University of California, Los Angeles, Los Angeles, CA.

Zymet, J. (2018). *Lexical propensities in phonology: corpus and experimental evidence from Slovenian and French* (Ph.D. Dissertation). University of California, Los Angeles, Los Angeles, CA.