Segmental Blocking in Dissimilation:
An Argument for Co-Occurrence Constraints

Juliet Stanton
Massachusetts Institute of Technology

1 Introduction

In processes of long-distance consonant dissimilation, one consonant is altered so as to avoid identity with another. A well-known example comes from [+lateral] dissimilation in Latin (Watkins 1970, Dressler 1971, Steriade 1987, Suzuki 1998, Cser 2010, Zymet 2015, Bennett 2015, *a.o.*). In Latin, the adjectival suffix /–ālis/ normally surfaces faithfully (1); if however the stem it attaches to contains an [l], then /–ālis/ can surface as /–āris/. In (2), /l...l/ maps to [l...r], reducing the similarity between the two segments.[1]

(1)　a.　/nāv+ālis/　→ [nāv-ālis]　　‘of ships, ship–, nautical, naval’

　　　b.　/hospit+ālis/　→ [hospit-ālis]　　‘of a guest, of a host, hospitable’

(2)　a.　/sōl+ālis/　→ [sōl-āris]　　‘of the sun, solar’

　　　b.　/lūk+ālis/　→ [lūk-āris]　　‘of or belonging to a grove’

What constraints motivate dissimilation? One popular answer is the OCP (Yip 1988, Suzuki 1998, cf. Bennett 2015). In Optimality Theory (Prince & Smolensky 2004), the OCP can be formalized as a set of constraints with the form *$X...X$ (esp. Suzuki 1998), where X denotes a feature or a feature matrix. For example, (2) can be analyzed under a ranking in which *[+lateral]...[+lateral] (3) dominates IO-IDENT[±lateral] (4).

(3)　*[+lateral]...[+lateral]: assign one violation mark for each pair of [+lateral] segments.

(4)　IO-IDENT[±lateral]: assign one violation for each [αlateral] output segment whose input correspondent is [-αlateral].

When /–ālis/ is added to a stem that does not contain an [l], it surfaces as such (5). When /–ālis/ is added to a stem that contains an [l] (as in (2)), it surfaces as [–āris], to satisfy *[+lateral]...[+lateral] (6). (I assume throughout that candidates like [sōr-ālis], where the stem [l] dissimilates, are ruled out by IO-IDENT(root).)

(5)

/nāv+ālis/	*[+lateral]...[+lateral]	IO-IDENT[±lateral]
☞　a.　[nāv+ālis]		
b.　[nāv+āris]		*!

(6)

/sōl+ālis/	*[+lateral]...[+lateral]	IO-IDENT[±lateral]
a.　[sōl-ālis]	*!	
☞　b.　[sōl-āris]		*

This paper focuses on one problem for a theory of dissimilation that appeals to *$X...X$ constraints: the existence of segmental blocking effects (name due to Bennett 2015), where an otherwise regular dissimilatory process is blocked by the presence of some (but not all) kinds of segments. For example, Latin [+lateral] dissimilation is blocked when an [r] intervenes between the trigger (stem) and target (suffixal) [l]s (7).

(7)　a.　/plūr+ālis/　→ [plūr-ālis], *[plūr-āris]　　‘of the plural’

　　　b.　/sepulcr-ālis/　→ [sepulcr-ālis], *[sepulcr-āris]　　‘of a tomb, sepuchral’

---

[1]　All Latin data discussed here comes from searches through Perseus. Note also that [+lateral] dissimilation is to some extent lexically variable, a point that I largely do not treat here (though see Section 2.1 for some brief discussion).

The failure of dissimilation to apply in (7) is not predicted by the current analysis. Given the ranking *[+lateral]…[+lateral] ≫ IO-IDENT[±lateral], we would not expect for the application of dissimilation to be sensitive to the identity of the material that intervenes between the two [+lateral] segments (8).

(8)

/plūr+ālis/	*[+lateral]…[+lateral]	IO-IDENT[±lateral]
☹ a. [plūr-ālis]	*!	
☞ b. [plūr-āris]		*

But as originally shown by Kenstowicz (1994), (7) can be analyzed by appealing to an additional co-occurrence constraint that disprefers pairs of [+rhotic] segments (see also e.g. Steriade 1995, Walsh-Dickey 1997, Cser 2010, Bennett 2015). Given input /plūr+ālis/, for example, the preference for attested [plūr-ālis] (with two [l]s) over the unattested *[plūr-āris] (with two [r]s) can be explained if *[+rhotic]…[+rhotic] ≫ *[+lateral]…[+lateral] (9).

(9)

/plūr+ālis/	*[+rhotic]…[+rhotic]	*[+lateral]…[+lateral]	IO-IDENT[±lateral]
☞ a. [plūr-ālis]		*	
b. [plūr-āris]	*!		*

The primary goal of this paper is to show that an analysis under which segmental blocking arises as an effect of competing co-occurrence constraints is sufficient to account for the typology of segmental blocking in long-distance consonant dissimilation, and that the constraints proposed under this analysis have parallel effects in the lexica of the relevant languages. Following this, I show that the co-occurrence-based analysis proposed here makes more accurately restrictive predictions regarding the typology of segmental blocking than do existing alternatives.

2   The typology of segmental blocking

This section provides an analysis of the three clear cases of segmental blocking in long-distance consonant dissimilation, identified from Bennett's (2015) comprehensive survey. These patterns (in Latin, Section 2.1; Georgian, Section 2.2; and Yidiɲ, Section 2.3), which constitute the total typology of blocking in dissimilation[2], can be analyzed by appealing to competing co-occurrence constraints.

2.1 *Latin [+lateral] dissimilation*   The analysis of Latin [+lateral] dissimilation previewed in the introduction can account for three generalizations: (i) when /–ālis/ attaches to a stem that does not contain an [l], it surfaces as [-ālis] (see (5)); (ii) when /–ālis/ attaches to a stem that contains an [l], it can surface as [-āris] (see (6)); (iii) unless an [r] intervenes between the two [l]s, in which case the suffix is [-ālis] (see (8)).[3] There is an additional generalization that has not yet been accounted for: in words where [r] precedes both [l]s, [+lateral] dissimilation applies (10). The current analysis incorrectly predicts that it should not (11).

(10) a. /rēgul+ālis/   → [rēgul-āris], *[rēgul-ālis]   'of a bar'

b. /rōmul+ālis/   → [rōmul-āris], *[rōmul-ālis]   *n.g.*

(11)

/rēgul+ālis/	*[+rhotic]…[+rhotic]	*[+lateral]…[+lateral]	IO-IDENT[±lateral]
☞ a. [rēgul-ālis]		*	
☹ b. [rēgul-āris]	*!		*

No one ranking of co-occurrence constraints can predict the correct outputs for both /plūr+ālis/ and /rēgul+ālis/: while deriving [plūr-ālis] requires the ranking *[+lateral]…[+lateral] ≫ *[+rhotic]…[+rhotic], deriving [rēgul-āris] requires the opposite (see also McMullin 2016:86 on this paradox in Georgian).

---

[2]   See the appendix (available on the author's website) on four additional possible cases. In each case there is good reason to doubt that 'blocking in dissimilation' is the only possible characterization. For discussion see also Bennett (2015).

[3]   Note that I adopt the characterization of the data assumed by Steriade (1987) and others: only an intervening [r] blocks [+lateral] dissimilation. While recent work has claimed that dissimilation is also blocked by an intervening non-coronal (e.g. Cser 2010, Bennett 2015), this claim is not supported by a statistical analysis of all relevant data (Zymet 2014a).

To solve this problem, the crucial observation is that *distance matters*. Note that for all Latin stems that contain both [l] and [r], suffixing /–ālis/ means that one of *[+lateral]...[+lateral] or *[+rhotic]...[+rhotic] must be violated. Given the input /plūr+ālis/, for example, the possible candidates contain either two [l]s (as in [plūr-ālis]) or two [r]s (as in [plūr-āris]). As shown in (12), the attested form is the one in which the co-occurring liquids are further away from one another, or less local. The input /hōrologi+ālis/ is included to show that the relevant metric of locality here is not just 'transvocalic vs. less local' (as has been claimed for other phenomena, e.g. McMullin & Hansson 2016); for the purposes of determining which co-occurring pair is less marked, the difference between 2 and 3 syllables is as important as the difference between 0 and 1.

(12)

Input	Attested	Unattested
/plūr+ālis/	[plūr-ālis] *[+lateral]...[+lateral] violated (1 syllable intervening)	[plūr-āris] *[+rhotic]...[+rhotic] violated (0 syllables intervening)
/rēgul+ālis/	[rēgul-āris] *[+rhotic]...[+rhotic] violated (2 syllables intervening)	[rēgul-ālis] *[+lateral]...[+lateral] violated (0 syllables intervening)
/hōrologi+ālis/	[hōrologi-āris] *[+rhotic]...[+rhotic] violated (3 syllables intervening)	[hōrologi-ālis] *[+lateral]...[+lateral] violated (2 syllables intervening)

The idea that the strength of a co-occurrence restriction depends on the distance between the offending segments is not new, and receives independent support from other facts about Latin [+lateral] dissimilation and the broader typology of long-distance dissimilation. While it has been tacitly assumed up until this point that Latin [+lateral] dissimilation applies if the appropriate conditions are met, the process is in fact lexically variable: the further apart the target and trigger [l]s, the less frequently dissimilation applies (see Zymet 2015; also Hurch 1991, Cser 2010). This Latin-specific fact (though see Zymet 2015 on similar effects in other languages) ties into a broader cross-linguistic generalization: if a dissimilatory process applies when the offending segments are separated by some distance *X*, it must also apply at all distances less than *X* (see esp. Suzuki 1998; cf. Bennett 2015). Recent work in artificial grammar learning suggests that this cross-linguistic generalization is psychologically real: participants have a difficult time learning patterns in which a dissimilatory alternation occurs in less-local (and not more-local) contexts (McMullin & Hansson 2016).

One way to formalize the notion that co-occurrence restrictions are sensitive to distance is to claim that co-occurrence constraints are gradiently evaluated (following Kimper 2011, Zymet 2015; cf. Martin 2005, Suzuki's 1998 Proximity Hierarchy), such that greater violations are assigned to more local co-occurring pairs. For example, *X...X* can be redefined gradiently as in (13). (There are many ways to implement the notion that the cost of violating *X...X* should increase as the segments draw closer together; see esp. Zymet 2015 for a Maxent alternative. The simple formulation in (13) is sufficient for the present purposes as the only crucial point here is that the smaller the distance between the two *X*s, the greater the violation of *X...X*.)

(13)   *X...X*: for each pair of co-occurring *X*s, assign...

    a. a violation of 5 if the pair occupies the same syllable,

    b. a violation of 4 if the pair occupies adjacent syllables,

    c. a violation of 3 if the pair is separated by one syllable,

    d. a violation of 2 if the pair is separated by two syllables, *etc.*

The analysis that follows is framed in Harmonic Grammar (Legendre et al. 1990). I make two assumptions here regarding violation assignment: categorical constraints assign a penalty of 1 for each locus of violation, and gradient constraints assign a penalty dependent on the violation (as in (13)). Violation costs for each constraint are determined by multiplying the penalty that the constraint assigns by its weight; the Harmony score associated with each candidate (represented as $\mathcal{H}$) is calculated by summing the violations it incurs.

In Latin, the fact that the more local violation of *[+lateral]...[+lateral] or *[+rhotic]...[+rhotic] is avoided can be explained by positing that the two constraints have the same weight. For example, an analysis under which both co-occurrence constraints have a weight of 2 (and assign penalties by (13)) and IO-IDENT[±lateral] has a weight of 1 (and always assigns a penalty of 1) predicts that the adjectival suffix /–ālis/ should always be realized as [-ālis] when the stem does not contain an [l] (14).

(14)

/nav+ālis/	*[+lateral]…[+lateral] *Weight: 2*	*[+rhotic]…[+rhotic] *Weight: 2*	IO-IDENT[±lateral] *Weight: 1*	$\mathcal{H}$
☞ a. [nav-ālis]				
b. [nav-āris]			-1	-1

If however an [l] is present in the stem, then /–ālis/ can be realized as [-āris] (15).

(15)

/sōl+ālis/	*[+lateral]…[+lateral] *Weight: 2*	*[+rhotic]…[+rhotic] *Weight: 2*	IO-IDENT[±lateral] *Weight: 1*	$\mathcal{H}$
a. [sōl-ālis]	-4			-8
☞ b. [sōl-āris]			-1	-1

If an [r] intervenes between the trigger and target [l]s, [+lateral] dissimilation is blocked, as (16b)'s more-local violation of *[+rhotic]…[+rhotic] is more costly than (16a)'s less-local violation of *[+lateral]…[+lateral].[4]

(16)

/plūr+ālis/	*[+lateral]…[+lateral] *Weight: 2*	*[+rhotic]…[+rhotic] *Weight: 2*	IO-IDENT[±lateral] *Weight: 1*	$\mathcal{H}$
☞ a. [plūr-ālis]	-3			-6
b. [plūr-āris]		-4	-1	-9

But if the [r] precedes the target and trigger [l]s, dissimilation applies normally: (17a)'s more-local violation of [+lateral]…[+lateral] is more costly than (17b)'s less-local violation of *[+rhotic]…[+rhotic].

(17)

/rēgul+ālis/	*[+lateral]…[+lateral] *Weight: 2*	*[+rhotic]…[+rhotic] *Weight: 2*	IO-IDENT[±lateral] *Weight: 1*	$\mathcal{H}$
a. [rēgul-ālis]	-4			-8
☞ b. [rēgul-āris]		-2	-1	-5

One potential challenge for this analysis comes from the observation that if [r] follows the target and trigger [l]s, [+lateral] dissimilation applies normally ((18); Suzuki 1998:106, following Walsh-Dickey 1997). These data cannot be accounted for by the above analysis: in the attested [salūtāriter] (18a), for example, the [r]s are in adjacent syllables. Had dissimilation failed to apply (yielding salūtāliter), the two [l]s would be separated by a syllable. For these forms, then, the present analysis predicts the wrong result.

(18) a. /salūt+āl+iter/ → [salūtāriter], *[salūtāliter] 'profitably, beneficially, advantageously'

  b. /singul+āl+iter/ → [singulāriter], *[singulāliter] 'particularly, exceedingly'

These data are unproblematic, however, if we assume that what motivates the appearance of the intermediate [r] in (18) is cyclic preservation of an [–āris] allomorph, present in the forms' morphological bases. Both [salūtāriter] and [singulāriter] have [–āris] bases, [salūtaris] and [singulāris]. If we assume that [salūtāriter] and [singulāriter] are formed by suffixing /–iter/ to the stems [salūtāri–] and [singulāri–], the appearance of the intermediate [r] in (18) can be attributed to a drive for base-derivative identity (e.g. Benua 1997).

The point here is that a theory of dissimilation referring only to *X…X constraints can account for the generalization that an intervening [r] blocks Latin [+lateral] dissimilation. The only necessary modification to the analysis is the independently motivated assumption that *X…X constraints are distance-scaled.

2.2 *Georgian [+rhotic] dissimilation* Georgian [+rhotic] dissimilation (Fallon 1993, Bennett 2015) is identical in nearly all relevant respects to Latin [+lateral] dissimilation, and is treated only briefly here. Certain [r]-containing suffixes alternate with [l]-containing allomorphs; I focus on the adjective forming /–uri/ (though see Fallon 1993:107–108 for others). As shown below, if there is no [r] in the stem, /–uri/ is realized faithfully as [-uri] (19); if however, the stem contains an [r], then /–uri/ can be realized as [-uli] (20).

---

[4] For forms like /largition+alis/, where the stem [l] and [r] are in the same syllable, the penalty that *[+lateral]…[+lateral] assigns to the attested [largition-alis] is equal to the penalty that *[+rhotic]…[+rhotic] assigns to unattested [largition-aris]. The fact that [largition-alis] is preferred is predicted by the presence of IO-IDENT[±lateral].

(19) a. /svan+uri/   → [svan-uri]   'Svan'

      b. /polon'+uri/   → [polon'-uri]   'Polish'

(20) a. /asur+uri/   → [asur-uli]   'Assyrian'

      b. /ast'ronomia+uri/   → [ast'ronomia-uli] 'astronomical'

Similarly to the Latin pattern, the ability of Georgian [l] to block [+rhotic] dissimilation depends on its position with respect to the target and trigger [r]s. If the [l] intervenes between the two [r]s, then [+rhotic] dissimilation is blocked (21). But if it precedes the two [r]s, [+rhotic] dissimilation applies normally (22).

(21) a. /ast'ral+uri/   → [ast'ral-uri]   'astral, of the stars'

      b. /parlament+uri/   → [parlament-uri]   'parliamentary'

(22) a. /bulgar+uri/   → [bulgar-uli]   'Bulgarian'

      b. /p'arlament'ar+uri/ → [p'arlament'ar-uli] 'parliamentary'

Georgian [+rhotic] dissimilation is the mirror image of Latin [+lateral] dissimilation: the most local violation of *[+lateral]…[+lateral] or *[+rhotic]…[+rhotic] is avoided. Examples follow in (23).

(23)

Input	Attested	Unattested
/ast'ral+uri/	[ast'ral-uri] *[+rhotic]…[+rhotic] violated (1 syllable intervening)	[ast'ral-uli] *[+lateral]…[+lateral] violated (0 syllables intervening)
/parlament+uri/	[parlament-uri] *[+rhotic]…[+rhotic] violated (3 syllables intervening)	[parlament-uli] *[+lateral]…[+lateral] violated (2 syllables intervening)
/bulgar+uri/	[bulgar-uli] *[+lateral]…[+lateral] violated (2 syllables intervening)	[bulgar-uri] *[+rhotic]…[+rhotic] violated (0 syllables intervening)

The analysis of Latin proposed above works without modification for the Georgian data presented here, as well as all additional data in Fallon 1993. But this result raises a question: if Georgian and Latin submit to the same analysis, why do the surface patterns look so different? Under this analysis, what differentiates the two is the default form of the alternating suffixes. In Latin, the default form of /–ālis/ (and the other alternating suffix, /–āl/) has an [l]. In Georgian, the default form of /–uri/ (and all other alternating suffixes) has an [r]. The fact that we see [+lateral] dissimilation in Latin but [+rhotic] dissimilation in Georgian, then, is due to a difference in the structure of their lexica – not a difference in the structure of their grammars.

2.3 *Yidiɲ [+lateral] dissimilation* In Yidiɲ, [+lateral] dissimilation (described in Dixon 1977:98–100) can be observed in certain multiple suffixation environments. For roots that belong to the [l]-conjugation class, what Dixon (1977) terms the 'going' aspect has the form /–ːliɲ/ (24). When the verbal comitative suffix [-ŋ-al] is added, /–ːliɲ/ can surface as [-ːri] (25).

(24) a. [magi-l]   → [magi-ːliɲ-u]   'climb up → went climbing up'

      b. [burwa-l]   → [burwa-ːliɲ-u]   'jump up → went jumping up'

(25) a. [ɖuŋgaɲa-liːɲ] → [ɖuŋgaɲa-riː-ŋal] 'went running (with)'

      b. [magi-ːliɲ-u]   → [magi-ːri-ŋal]   'went climbing up (with)'

For some derivatives, however, [+lateral] dissimilation fails. Dixon (1977:100) notes that when /–ːliɲ/ and /–ŋal/ are added to roots containing a rhotic (either alveolar [r] or retroflex [ɻ]), both surface faithfully (26). From this, we can infer that the presence of a rhotic blocks [+lateral] dissimilation.

(26) a. [burwa-l]   → [burwa-ːli-ŋal] 'jump → went jumping with'

      b. [burgin-n]   → [burgi-ːli-ŋal] '(went) walk about → (went) walk about with'

In Yidiɲ, unlike Latin and Georgian, the status of a consonant as a blocker is not crucially tied to its status as an intervener: as is evident from (26), a rhotic does not have to intervene between the target and trigger [l]s in order to block [+lateral] dissimilation. This reveals a difference in priorities. In both Latin and Georgian, whether it is better to violate *[+lateral]...[+lateral] or *[+rhotic]...[+rhotic] depends on the locality of the co-occurring pair. In Yidiɲ, however, avoiding violation of *[+rhotic]...[+rhotic] is always more important than avoiding violation of *[+lateral]...[+lateral] (Crowhurst & Hewitt 1995:79, Steriade 1995:154, Walsh-Dickey 1997:164, Suzuki 1998:107ff).

To derive the Yidiɲ pattern, in which words with multiple [l]s are always preferred, it must be the case that *[+lateral]...[+lateral] has a lower weight than *[+rhotic]...[+rhotic]; the tableaux below assume that their weights are 2 and 10, respectively. As shown in (27), when there is no rhotic in the stem, the analysis correctly predicts that [+lateral] dissimilation should occur.

(27)

/magi+ːli+ŋal/	*[+rhotic]...[+rhotic] *Weight: 10*	*[+lateral]...[+lateral] *Weight: 2*	IO-IDENT[±lat] *Weight: 1*	ℋ
a. [magi-ːli-ŋal]		-4		-8
☞ b. [magi-ːri-ŋal]			-1	-1

When there is a rhotic in the stem, however, [+lateral] dissimilation does not occur: even though (28a)'s violation of *[+lateral]...[+lateral] is more local, (28b)'s violation of *[+rhotic]...[+rhotic] is more costly.

(28)

/burwa+ːli+ŋal/	*[+rhotic]...[+rhotic] *Weight: 10*	*[+lateral]...[+lateral] *Weight: 2*	IO-IDENT[±lat] *Weight: 1*	ℋ
☞ a. [burwa-ːli-ŋal]		-4		-8
b. [burwa-ːri-ŋal]	-3		-1	-31

3 Evidence from the lexicon

If the analyses presented in Section 2 are correct, they require us to claim that distance-scaled versions of both *[+lateral]...[+lateral] and *[+rhotic]...[+rhotic] are active in the phonologies of Latin, Georgian, and Yidiɲ. This section presents preliminary evidence that these constraints can also be linked to trends in the lexica of the three relevant languages.

3.1 *Methods*   One method commonly used to assess whether or not two consonants co-occur as often as expected is the O/E (Observed/Expected) ratio (Pierrehumbert 1992, *a.o.*). The O/E ratio is used to compare the number of forms in a corpus that exhibit some property *P* to the number of forms that we would expect to exhibit *P*, were there no restrictions on *P*'s occurrence. If fewer forms that exhibit *P* are observed than are expected to occur, this indicates that forms exhibiting *P* are underattested, relative to a naïve expectation.

The O/E ratios in this section were calculated with a Python script (available from the author upon request) that took three factors into account. First, different consonants occur at different frequencies in the lexicon: in Latin, for example, [r] is less common than [t]. Second, the expected frequency of a given consonant varies by context: rhotics in Yidiɲ, for example, do not occur word-initially. And third, the expected frequency at which two consonants co-occur depends on the length of the word: the longer the word, the more chances the consonant has to appear.

3.2 *Results: Latin*   Preliminary evidence that both *[+lateral]...[+lateral] and *[+rhotic]...[+rhotic] have observable effects in the Latin lexicon comes from Cser (2010). Regarding [l], Cser (p. 35) writes that "multiple occurrences of [...] l within stems are rare"; regarding [r], Cser (p. 42) writes that "irrespective of morphological composition, the minimum distance between two r's is a long vowel or two segments except in final syllables where a short vowel is enough".

To see whether or not these restrictions on words containing multiple [l]s and [r]s are more severe than we might expect, I analyzed a wordlist containing 1,732 nominative singular nouns.[5] The results are in (29); statistical significance was assessed via the Fisher's Exact test (is observed/expected significantly different than expected/expected?). In both cases, O/E < 1, though in the case of [r...r] this is not significant.

---

[5]   Thanks to Adam Albright for the Latin wordlist.

(29)

Sequence	Observed	Expected	O/E	Significant?
[l...l]	10 forms	43 forms	0.2	Yes ($p < .001$)
[r...r]	69 forms	103 forms	0.7	No ($p = .06$)

The results in (29) are consistent with the claim that both *[+lateral]...[+lateral] and *[+rhotic]...[+rhotic] are active in Latin. Perhaps more important, though, is the observation that the lexical underattestation patterns, like the categorical dissimilatory alternation, appear to display distance effects. Of the 10 words in the wordlist that contain two or more [l]s, in all but one (*līlium*), the [l]s are separated by at least one syllable (e.g. *lectulus*, *adulescentulus*). Of the 69 words that contain two or more [r]s, in 66 the two [r]s are separated by at least two segments (i.e. *arbor*); in the remaining three, the intervening vowel is either long (as in *prōra*, *aurōra*) or in the final syllable (*furor*; see Cser 2010:43ff on *-or/-ōris* nouns).

These results indicate that the restrictions found in the lexicon parallel the restrictions that drive [+lateral] dissimilation. Regarding the effects of *[+lateral]...[+lateral]: Zymet (2015) shows that [+lateral] dissimilation applies categorically when the two [l]s are in adjacent syllables. This effect has parallels in the lexicon, where we find a near-categorical ban on words that have two [l]s in adjacent syllables. Regarding the effects of *[+rhotic]...[+rhotic]: in the lexicon, there is a near-categorical ban on words that contain [r]s in adjacent syllables. This has parallels in the behavior of [r] in [+lateral] dissimilation: of the 20 cases in Perseus where [r] directly precedes /–ālis/ (e.g. in [plūr-ālis]), in 19 the [r] blocks dissimilation. These links between [+lateral] dissimilation, [+rhotic] blocking, and trends in the lexicon provide support for an analysis that penalizes the co-occurrence of multiple [l]s and multiple [r]s within a word. The explanation of these links, under the present analysis, is that the same constraints that drive the dissimilatory suffixal alternations (*[+rhotic]...[+rhotic] and *[+lateral]...[+lateral]) also shape the lexicon.

Before moving on, a note is necessary about the generality of the effects of *[+lateral]...[+lateral] in Latin. While Cser's claims regarding the restrictions on words containing multiple [r]s hold throughout the derived lexicon, Latin has other [l]-containing suffixes which do not undergo dissimilation when attached to an [l]-containing stem; examples from Cser 2010:40–41 are in (30).

(30)   a. *-culus/-cula/-culum* (e.g. *folliculus* 'bag', *plebecula* 'mob')

b. *-bulum* (e.g. *latibulum* 'den')

c. *-bilis/-bile* (e.g. *tolerabilis* 'tolerable')

d. *-ilis* (e.g. *pluviatilis* 'rain-', *electilis* 'select')

e. diminutive *-l* (e.g. *ollula* 'jug', *villula* 'mansion')

There are several available ways of explaining why /–ālis/ undergoes dissimilation, but the suffixes in (30) do not. One is to claim that /–ālis/ has two listed allomorphs (one [l]-containing, one [r]-containing) and that the suffixes in (30) differ in that they have only one. Another option is to claim that the faithfulness constraints that govern the realization of the suffixes in (30) take priority over *[+lateral]...[+lateral], but the faithfulness constraints for /–ālis/ do not. (There is, in addition, a question here about whether or not *[+lateral]...[+lateral] can inhibit derivation: do the suffixes in (30) attach to [l]-containing stems less often than we would expect for them to? While there is some suggestive evidence that the answer is yes – of the 210 *-ilis* derivatives in Perseus, none has an [l]-final stem – more investigation is necessary on this point.)

3.3   *Results: Georgian and Yidiɲ* While the lexica available for Georgian and Yidiɲ are much smaller than the one available for Latin, the results in both cases are largely consistent with the Latin results.

For Georgian, I analyzed a wordlist of 940 words included in Vogt's (1971) grammar.[6] The results are presented in (31); while the observed and expected counts for words containing multiple [l]s are perhaps too low to conclude anything from, it is clear that there exists a strong dispreference for words that contain multiple [r]s (see also Fallon 1993:111, who claims that this dispreference results in a categorical ban).

(31)

Sequence	Observed	Expected	O/E	Significant?
[l...l]	4 forms	5 forms	0.8	No ($p = 1$)
[r...r]	5 forms	25 forms	0.2	Yes ($p < .001$)

---

[6]   The wordlist includes 1069 words, but 129 were excluded as they did not have vowels.

The Georgian results, too, are consistent with the claim that co-occurrence constraints are distance-scaled: in the 5 words with multiple [r]s, in 4 the [r]s are maximally distant (*ragvar*, *ranair*, *rogor*, *rogor*; cf. *rarig*).

For Yidiɲ, I analyzed a wordlist of 920 words included in Dixon's (1977) grammar. As is evident from (32), while most of the O/E ratios are under 1, there is generally not enough data to allow us to determine if these results diagnose any co-occurrence restrictions.

(32)

Sequence	Observed	Expected	O/E	Significant?
[l…l]	16 forms	21 forms	0.8	No ($p = .6$)
[r…r]	8 forms	9 forms	0.9	No ($p = 1$)
[ɽ…ɽ]	9 forms	7 forms	1.3	No ($p = 1$)
[r/ɽ…r/ɽ]	6 forms	15 forms	0.4	No ($p = .2$)

It is fairly clear, however, that Yidiɲ exhibits a marked dispreference for words that contain multiple [r]s. Each of the 8 forms that contains multiple [r]s in Dixon's wordlist is reduplicated (also Suzuki 1998:110), which suggests that there is a categorical ban on the co-occurrence of [r] within a single Yidiɲ stem. As far as I am aware, this ban is completely general: I have been unable to find forms in Dixon's grammar in which an [r]-containing suffix attaches to an [r]-containing stem.[7] This effect is consistent with the existence of highly-weighted *[+rhotic]…[+rhotic], necessary to account for blocking in Yidiɲ [+lateral] dissimilation.

4   Comparison with alternatives

The goal of this paper thus far has been to demonstrate that segmental blocking can be analyzed under the assumption that all dissimilatory effects are motivated by *X…X constraints. There are, however, other theories of dissimilation that acknowledge the existence of either other constraints or additional conditions on *X…X constraints. This section considers the predictions of two alternatives and argues that they are less accurately restrictive than the predictions of the co-occurrence-based analysis outlined above.

4.1   *Tier-based locality*   An alternative account of the blocking patterns in Latin and Georgian appeals to tier-based locality. For example, Steriade's (1987) analysis of Latin [+lateral] dissimilation claims that the driving force between the alternation is a ban on tier-adjacent [+lateral] segments (i.e. those that are not separated by a [-lateral] [l]; this analysis relies on the assumption that all other segments are not specified for [±lateral]). Thus [+lateral] dissimilation applies in /sōl+ālis/ and /rēgul-ālis/ because there is no [r] that intervenes between the two [+lateral] [l]s; dissimilation is however blocked in /plūr+ālis/ because there is an intervening [r]. (For a similar analysis of Georgian [+rhotic] dissimilation, see Fallon 1993; also McMullin 2016 on deriving blocking in Georgian and elsewhere with tier-based locality.)

As noted by Steriade (1995:154), however, this analysis does not generalize to Yidiɲ, where [+lateral] dissimilation is blocked by a non-intervening [r]. Without further modification, the account of blocking that appeals to tier-based locality undergenerates, as it cannot capture the full typology.

4.2   *Dissimilation as surface correspondence*   In the Surface Correspondence Theory of Dissimilation (SCTD; Bennett 2015), dissimilatory processes are analyzed as an interaction between two constraint families. The first is a family of CORR[F] constraints, which require consonants that share some feature or feature matrix [F] to correspond. CORR[+lateral], for example, demands that all [+lateral] consonants in a given domain must correspond ((33); correspondence relations are indicated with subscripts).

(33)

/sōl+ālis/	CORR[+lateral]
☞ a. sōl$_x$-āl$_x$is	
b. sōl$_x$-āl$_y$is	*!

The second family is made up of CC-Limiter constraints, which place requirements on corresponding consonants. CC-EDGE(root), for example, requires that corresponding consonants reside on the same side of a root boundary (i.e. all corresponding consonants may be in or outside the root, but it is not possible for one correspondent to be in the root and the others outside). The activity of CC-EDGE(root) is illustrated in (34).

---

[7]   In fact, there appears to be only one [r]-containing suffix (an allomorph of the imperative marker) that resides within the stem domain. There are other [r]-containing suffixes that do not belong to the same phonological word as the stem they attach to (Dixon 1977:90).

(34)

/sōl+ālis/	CC-EDGE(root)
a. sōl$_x$-āl$_x$is	*!
☞ b. sōl$_x$-āl$_y$is	

Under this theory, dissimilation arises when correspondence is both required (by a CORR constraint) and prohibited (by a CC-Limiter constraint). In (35), for example, the candidate in which the two [l]s correspond across a root boundary violates CC-EDGE(root); the candidate in which the two [l]s do not correspond violates CORR[+lateral]. The optimal candidate then is (35c), in which dissimilation occurs (violating only low-ranked IO-IDENT[±lateral]) and both CORR[+lateral] and CC-EDGE(root) are vacuously satisfied.

(35)

/sōl+ālis/	CORR[+lateral]	CC-EDGE(root)	IO-IDENT[±lateral]
a. sōl$_x$-āl$_x$is		*!	
b. sōl$_x$-āl$_y$is	*!		
☞ c. sōl$_x$-ār$_y$is			*

This subsection explores an analysis of blocking in the SCTD, building on the analyses of Latin, Georgian, and Yidiɲ proposed by Bennett (2015: Ch. 8). I show first that this theory undergenerates: it cannot predict that the status of a consonant as a blocker should depend on its position with respect to the trigger and target segments, and it does not capture parallels between the categorical suffixal alternations in all three languages and co-existing co-occurrence restrictions within roots. These problems can be fixed, but the necessary modifications cause the theory to overgenerate. The main point is that the STCD provides a less restrictive fit to the typology of segmental blocking than does the co-occurrence based analysis.

4.2.1 *The locality problem* The first overgeneration problem faced by the SCTD is that, in its current form, it cannot account for the fact that a consonant must intervene to block dissimilation in Latin and Georgian. To see why this problem arises, consider how an analysis of the Latin data might proceed.[8] The fact that [r] blocks [+lateral] dissimilation can be analyzed as an interaction between two CORR constraints: CORR[+rhotic] and CORR[+lateral]. As shown in (36), if CORR[+rhotic] ≫ CORR[+lateral], a candidate with non-corresponding [l]s (36b) is preferable to a candidate with non-corresponding [r]s (36d).

(36)

/plūr+ālis/	CORR[+rhotic]	CC-EDGE(root)	CORR[+lateral]
a. [pl$_x$ūr$_y$-āl$_x$is]		*!	
☞ b. [pl$_x$ūr$_y$-āl$_z$is]			*
c. [pl$_x$ūr$_y$-ār$_y$is]		*!	
d. [pl$_x$ūr$_y$-ār$_z$is]	*!		

This analysis also makes the incorrect prediction, however, that a non-intervening [r] should also block [+lateral] dissimilation: given the ranking CORR[+rhotic] ≫ CORR[+lateral], it should always be the case that a form with two [l]s is preferred to a form with two [r]s. While this may be the correct analysis of the Yidiɲ pattern, it is not the result that we want for Latin.

(37)

/rēgul+ālis	CORR[+rhotic]	CC-EDGE(root)	CORR[+lateral]
a. [r$_x$ēgul$_y$-āl$_y$is]		*!	
💣 b. [r$_x$ēgul$_y$-āl$_z$is]			*
☹ c. [r$_x$ēgul$_y$-ār$_x$is]		*!	
☹ d. [r$_x$ēgul$_y$-ār$_z$is]	*!		

One solution to this problem, following Zymet (2014b), is to claim that CORR constraints are distance-sensitive: the closer together a pair of consonants, the more pressure for them to correspond. For one possible redefinition along these lines, see (38). With this modification, the SCTD is capable of accounting for the known typology of blocking effects in dissimilation. As for the OCP-based analysis in Section 2.2, to derive the Georgian and Latin facts, locality effects emerge when CORR[+rhotic] and CORR[+lateral] have the same weight. To derive the Yidiɲ facts, CORR[+rhotic] must be weighted higher than CORR[+lateral].

---

[8] Bennett's actual analysis of the Latin data differs, as he assumes a different characterization of the data. See fn. 3.

(38)   CORR[F](DOM): all consonants bearing [F] that reside within some domain DOM must correspond. For each pair that does not, assign. . .

    a. a score of 5 if the pair occupies the same syllable,

    b. a score of 4 if the pair occupies adjacent syllables, *etc.*

4.2.2 *The root vs. affix problem*   Recall from Section 2.4 that, in the three languages under discussion here, the restrictions on laterals and rhotics that appear to motivate both dissimilation and blocking hold not just across a suffix boundary, but also within roots. The current correspondence-based analysis, however, does not predict that the restrictions should hold root-internally.

    In the analysis of the Latin facts developed above, dissimilation is compelled in part by the impossibility of correspondence across a root boundary. If two [l]s co-occur within the root, the current analysis predicts that they should correspond and co-exist (as in *lilium*, (39)).

(39)

/lilium/		CC-EDGE(root)	CORR[+lateral]	IO-ID[±lateral]
☞ a.	$[l_x il_x ium]$			
b.	$[l_x il_y ium]$		*!	
c.	$[l_x ir_x ium]$			*!
d.	$[l_x ir_y ium]$			*!

In order to capture the fact that co-occurring [l]s are dispreferred within roots, it is necessary to replace CC-EDGE(root) with two different CC-Limiter constraints: CC-SROLE (40) and CC-EDGE(σ) (41).

(40)   CC-SROLE: corresponding consonants must have matching syllabic roles.

(41)   CC-EDGE(σ): corresponding consonants must belong to the same syllable.

Together, these constraints ban correspondence between consonants that do not occupy the same syllabic role in the same syllable (as illustrated in (42–43)); in other words, any pair of distinct consonants.

(42)

/lala/		CC-EDGE(σ)	CC-SROLE	CORR[+lateral]	IO-ID[±lateral]
a.	$[l_x al_x a]$	*!			
b.	$[l_x al_y a]$			*!	
☞ c.	$[l_x ar_y a]$				*

(43)

/lal/		CC-EDGE(σ)	CC-SROLE	CORR[+lateral]	IO-ID[±lateral]
a.	$[l_x al_x]$		*!		
b.	$[l_x al_y]$			*!	
☞ c.	$[l_x ar_y]$				

As noted by McMullin & Hansson (2016), the only known way to motivate context-free dissimilation in the SCTD is with the constraints in (40–41). With this modification, the SCTD is capable of accounting for the suffixal alternations as well as the root-internal restrictions present in the three cases documented here.

4.3 *Comparison*   As is clear from the above discussion, the co-occurrence- and correspondence-based theories of dissimilation are both capable of accounting for the extant typology of segmental blocking. This subsection asks, through factorial typology, if the predictions of the co-occurrence-based theory differ from those of the correspondence-based theory. The preliminary result reported here is that the predictions do diverge, and that the co-occurrence-based theory is more restrictive.

4.3.1 *Methods*   The constraint sets used to explore the factorial typology for each theory were those constraints necessary to analyze the typology of segmental blocking – both the alternations and their associated effects in the lexicon. For the co-occurrence-based theory, these constraints are *[+rhotic]. . .[+rhotic], *[+lateral]. . .[+lateral], and IO-IDENT[±lateral]; for the correspondence-based theory, these are CORR[+rhotic], CORR[+lateral], CC-SROLE, CC-EDGE(σ), and IO-IDENT[±lateral]. All constraints are defined elsewhere in this paper; the candidate sets are available in the appendix (available at the author's website).

The factorial typologies were explored in OTSoft (Hayes et al. 2013), under the assumption that constraints stand in relationships of strict domination. Both constraints sets thus necessarily undergenerate, as the assumption of strict domination means that Yidiɲ-like patterns (where one co-occurrence constraint consistently takes precedence over another) are possible but Latin/Georgian-type patterns (where locality is crucial to determining which co-occurrence constraint takes precedence) are not. Effectively, this means that the current results represent only a fraction of the predicted typological space.

4.3.2 *Results*   The co-occurrence-based constraint set predicted five different patterns, summarized in (44). Of these five, four are attested.

(44)   Results of co-occurrence-based factorial typology

   a. Description: combinations of [r]s and [l]s surface faithfully
      Ranking: IO-IDENT[±lateral] ≫ *[+rhotic]...[+rhotic], *[+lateral]...[+lateral]
      Examples: many

   b. Description: combinations of [l]s surface faithfully. Combinations of [r]s are banned.
      Ranking: *[+rhotic]...[+rhotic] ≫ IO-IDENT[±lateral] ≫ *[+lateral]...[+lateral]
      Example: Kammu (Svantesson 1983, Bennett 2015)

   c. Description: combinations of [r]s surface faithfully. Combinations of [l]s are banned.
      Ranking: *[+rhotic]...[+rhotic] ≫ IO-IDENT[±lateral] ≫ *[+lateral]...[+lateral]
      Example: Persian (Kambuziya et al. 2009, Bennett 2015)

   d. Description: combinations of [r]s and [l]s are banned. If a word must have either two [r]s or two [l]s, the form with two [r]s is chosen.
      Ranking: *[+lateral]...[+lateral] ≫ *[+rhotic]...[+rhotic] ≫ IO-IDENT[±lateral]
      Example: —

   e. Description: combinations of [r]s and [l]s are banned. If a word must have either two [l]s or two [r]s, the form with two [l]s is chosen.
      Ranking: *[+rhotic]...[+rhotic] ≫ *[+lateral]...[+lateral] ≫ IO-IDENT[±lateral]
      Example: Yidiɲ (Dixon 1977, *a.o.*)

The correspondence-based constraint set predicts all patterns in (44), as well as 15 additional distinct patterns. In some of these additional systems, liquids are only allowed to co-occur if they share the same syllabic role; in others (as in (45)), liquids are only allowed to co-occur if they occupy the same syllable. The prediction that dissimilation should only apply outside a syllable is, as far as I know, unsupported (*pace* Bennett 2015), and in some cases the reverse appears to hold (see e.g. Martin 2005:20–27 on liquid dissimilation in English).

(45)   Pattern: transsyllabic [r]-dissimilation
       Ranking: CORR[+rhotic], CC-EDGE(σ) ≫ IO-IDENT[±lateral] ≫ CORR[+lateral] ≫ CC-SROLE

   a. [r]-dissimilation if the [r]s do not reside within the same syllable
      (i) /rara/   → [rala]
      (ii) /ralda-ra/ → [ralda-ra]
   b. No dissimilation if the [r]s do reside within the same syllable
      (i) /rar/   → [ral]
      (ii) /lar+ar/   → [lar-ar]

In the rest of the additional patterns uniquely predicted by the correspondence-based approach, concurrent restrictions on [r]s and [l]s manifest themselves in different structural configurations (in one case, for example, [r]s are banned unless they are in the same syllable and [l]s are banned if they do not share a syllabic role). All 15 of the systems uniquely predicted by the correspondence-based constraint set do not resemble any known patterns; in other words, the correspondence-based theory overgenerates, relative to the co-occurrence-based one. The remainder of the predicted patterns and their analysis are in the online appendix.

While it is necessary to reiterate at this point that the results presented here only represent a fraction of the predicted typology, preliminary results indicate that allowing the constraint set to reference gradient locality only serves to widen the gap between the number of systems predicted by the co-occurrence-based and the correspondence-based analyses. I leave a fuller exploration of the factorial typologies to future work.

5   Conclusion

In this paper I have argued that an co-occurrence-based theory of dissimilation makes more accurately restrictive predictions than competing alternatives regarding the typology of segmental blocking in consonant dissimilation. This is a point in its favor.

References

Bennett, Wm. G. (2015). *The Phonology of Consonants: Harmony, Dissimilation, and Correspondence*. Cambridge University Press, Cambridge.

Benua, Laura (1997). *Transderivational Identity: Phonological Relations Between Words*. Ph.D. thesis, UMass Amherst, Amherst, MA.

Crowhurst, Megan & Mark Hewitt (1995). Prosodic Overlay and Headless Feet in Yidiny. *Phonology* 12, 39–84.

Cser, András (2010). The-alis/aris-allomorphy revisited. Rainer, Franz, Wolfgang U. Dressler, Dieter Kastovsky & Hans C. Luschützky (eds.), *Variation and change in morphology: selected papers from the 13th international morphology meeting*, John Benjamins, Amsterdam/Philadelphia, 33–51.

Dixon, Robert M. W. (1977). *A Grammar of Yidin*. Cambridge University Press, Cambridge.

Dressler, Wolfgang (1971). An Alleged Case of Non-chronological Rule Insertion. *Linguistic Inquiry* 2, 597–599.

Fallon, Paul D. (1993). Liquid dissimilation in Georgian. Kathol, Andreas & Michael Bernstein (eds.), *Proceedings of the 10th Eastern States Conference on Linguistics (ESCOL)*, DMLL Publications, Ithaca, NY, 105–116.

Hayes, Bruce, Bruce Tesar & Kie Zuraw (2013). OTSoft 2.3.2, software package. http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Hurch, Bernard (1991). On Adjacency and Related Concepts. Bertinetto, Pier Marco, Michael Kenstowicz & Michele Loporcaro (eds.), *Certamen Phonologicum II*, Rosenberg and Sellier, Torino, 43–63.

Kambuziya, Aliyeh Kord-e Zafaranlu, Ferdos Aghagolzadeh, Arsalan Golfam & Bashir Jam (2009). An Optimality-Theoretic Account of Dissimilation in Persian. *The International Journal of Humanities* 2, 59–74.

Kenstowicz, Michael (1994). *Phonology in Generative Grammar*. Blackwell, Oxford.

Kimper, Wendell A. (2011). *Competing triggers: Transparency and opacity in vowel harmony*. Ph.D. thesis, University of Massachusetts, Amherst.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar - A Formal Multi-Level Connectionist Theory of Linguistic Well-Formedness: Theoretical Foundation. ICS Technical Report 90-5, University of Colorado, Boulder, CO.

Martin, Andrew Thomas (2005). *The Effects of Distance on Lexical Bias: Sibilant Harmony in Navajo Compounds*. Master's thesis, University of California, Los Angeles.

McMullin, Kevin & Gunnar Ólafur Hansson (2016). Long-Distance Phonotactics as Tier-Based Strictly 2-Local Languages. Albright, Adam & Michelle A. Fullwood (eds.), *Proceedings of the 2014 Meeting on Phonology*, Linguistic Society of America, Washington, DC.

McMullin, Kevin James (2016). *Tier-Based Locality in Long-Distance Phonotactics: Learnability and Typology*. Ph.D. thesis, The University of British Columbia, Vancouver, BC.

Pierrehumbert, Janet (1992). Dissimilarity in the Arabic Verbal Roots. *Proceedings of the 23rd Meeting of the Northeastern Linguistic Society*, GLSA, Amherst, MA.

Prince, Alan & Paul Smolensky (2004). *Optimality Theory: Constraint interaction in generative grammar*. Blackwell, Oxford.

Steriade, Donca (1987). Redundant values. Bosch, A., B. Need & E. Schiller (eds.), *Papers from the parasession on metrical and autosegmental phonology*, Chicago Linguistic Society, Chicago, 339–362.

Steriade, Donca (1995). Underspecification and Markedness. Goldsmith, John (ed.), *The Handbook of Phonological Theory*, Blackwell, Cambridge, MA, 114–174.

Suzuki, Keiichiro (1998). *A typological investigation of dissimilation*. Ph.D. thesis, The University of Arizona, Tucson.

Svantesson, Jan-Olof (1983). *Kammu Phonology and Morphology*. CWK Gleerup, Malmö.

Vogt, Hans (1971). *Grammaire de la langue géorgienne*. Universitetsforlaget, Oslo.

Walsh-Dickey, Laura (1997). *The Phonology of Liquids*. Ph.D. thesis, UMass Amherst, Amherst, MA.

Watkins, Calvert (1970). A case of non-chronological rule insertion. *Linguistic Inquiry* 1, 525–527.

Yip, Moira (1988). The OCP and phonological rules: a loss of identity. *Linguistic Inquiry* 19, 65–100.

Zymet, Jesse (2014a). *Distance-based decay in long-distance phonological processes: a probabilistic model for Malagasy, Latin, English, and Hungarian*. Master's thesis, University of California, Los Angeles.

Zymet, Jesse (2014b). Distanced-based decay in long-distance phonological processes. Poster presented at ABC↔Conference, Berkeley, CA.

Zymet, Jesse (2015). Distance-Based Decay in Long-Distance Phonological Processes. et al., Ulrike Steindel (ed.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics*, Cascadilla Proceedings Project, Somerville, MA, 72–81.