2014

Canadian Raising with Language-Specific Weighted Constraints

Joe Pater

Canadian raising with language-specific weighted constraints

Joe Pater, University of Massachusetts Amherst

The distribution of the raised variants of the Canadian English diphthongs is standardly analyzed as opaque allophony, with derivationally ordered processes of diphthong raising and of /t/ flapping. This paper provides an alternative positional contrast analysis in which the pre-flap raised diphthongs are licensed by a language-specific constraint. The basic distributional facts are captured with a weighted constraint grammar that lacks the intermediate level of representation of the standard analysis. The paper also provides a proposal for how the constraints are learned, and shows how correct weights can be found with a simple, widely used learning algorithm.[*]

1. Introduction

In Canadian English, the diphthongs [ai] and [au] are famously in near-complementary distribution with raised variants [ʌi] and [ʌu]. For the most part, the raised diphthongs occur only before tautosyllabic voiceless consonants (1a.), with their lower counterparts occurring elsewhere (1b.). The distribution overlaps only before the flap [ɾ] (1c.).

(1) a. [sʌik] *psych* [rʌit] *write* [lʌif] *life* [hʌus] *house*
b. [ai] *I* [raid] *ride* [laivz] *lives (pl.)* [hauz] *house (v.)*
c. [mʌiɾɚ] *mitre* [saiɾɚ] *cider* [tʌiɾl] *title* [braiɾl] *bridle*
[rʌiɾɚ] *writer* [raiɾɚ] *rider*

As Idsardi (2006) points out, analyses of CANADIAN RAISING (Chambers 1973) are generally of two types: those that treat the low/raised diphthong distinction as phonemic (Joos 1942), and those that treat it as opaquely allophonic, with the surface vowel contrast derived from the underlying contrast between /t/ and /d/ that is itself neutralized to the flap (Harris 1951/1960). The standard analysis is that of Chomsky 1964, in which the rule that raises underlying /ai/ and /au/ to [ʌi] and [ʌu] before voiceless consonants applies before the rule changing underlying /t/ to [ɾ], producing derivations like /taitl/ → tʌitl → [tʌiɾl].

In this paper, I pursue a third type of analysis, intermediate between the phonemic and allophonic approaches, in which the distribution of these diphthongs is an instance of positionally restricted contrast (see Mielke et al. 2003 for an earlier positional contrast analysis).[1]

---

[*] Thanks especially to Paul Boersma for his collaboration on an earlier presentation of this research (Boersma and Pater 2007), and to Michael Becker, Elliott Moreton, Jason Narad, Presley Pizzo and David Smith for their collaboration on versions of the constraint induction procedure described in section 3. Thanks also to Adam Albright, Eric Baković, Ricardo Bermúdez-Otero, Jack Chambers, Heather Goad, Bruce Hayes, Bill Idsardi, Karen Jesney, John McCarthy, Robert Staubs and Matt Wolf for useful discussion, as well as participants in Linguistics 751, UMass Amherst in 2008 and 2011, and at NELS 38, University of Ottawa, and MOT 2011, McGill University. This research was supported by grant BCS-0813829 from the National Science Foundation to the University of Massachusetts Amherst.

[1] Mielke et al. (2003) restrict the distribution of by disallowing voiced consonants after [ʌi] and [ʌu], and voiceless consonants after [ai] and [au]. This analysis fails to capture the ill-formedness of [ʌi] and [ʌu]

The first apparent challenge for this approach is that the environment for the contrast is phonetically unnatural: the flap has no known phonetic property that makes raised diphthongs easier to produce or perceive before it. This is a problem, of course, only if phonological rules or constraints are limited to those that are phonetically grounded. Such a limit has long been known to be untenable (Bach and Harms 1972, Anderson 1981). In particular, there are well-documented instances of productive phonological patterns that do not have a synchronic phonetic basis (see e.g. Icelandic velar fronting in Anderson 1981, NW Karaim consonant harmony in Hannson 2007, and Sardinian [l] ~ [ʁ] alternations in Scheer 2014; see also Hayes et al. 2009 for discussion and experimental work). Furthermore, experimental studies have found little, if any, evidence that phonetically grounded patterns enjoy a special status in learning, although other factors, such as structural simplicity, have a consistent effect (see Moreton and Pater 2012 for an overview). In the analysis to follow, pre-flap raised diphthongs are licensed by a language-specific phonetically arbitrary constraint.

A second potential challenge for an analysis of Canadian raising with positional contrast is to rule out raised diphthongs in environments other than those of a following voiceless obstruent or flap. I show that it is possible to properly restrict the distribution of raised diphthongs with a small set of constraints, if those constraints are weighted, as in HARMONIC GRAMMAR (HG; Smolensky and Legendre 2006; see Pater 2009, 2014, for an introduction and overview of other research in this framework).

This paper also includes a proposal for how the constraints in HG are learned. I adopt a broadly used on-line learning algorithm that Boersma and Pater (2014) refer to as the HG-GLA. In the proposed extension to constraint induction, constraints are constructed from differences between the structure of the observed forms and the learner's "mistakes". I illustrate this approach using a simplified version of the distribution of Canadian English diphthongs.

The analysis makes use of only a single mapping from underlying representation (UR) to surface representation (SR), with no intermediate derivational levels, as in standard optimality theory (OT: Prince and Smolensky 1993/2004). Some other basic assumptions differ from those of standard OT: the constraints are weighted, rather than ranked, and the constraints are language-specific and sometimes phonetically arbitrary, rather than universal and substantively grounded. This diverges from approaches to the distribution of the raised diphthongs and to other instances of 'counterbleeding opacity' that maintain OT's ranking and universality but enrich its derivational component (see Bermúdez-Otero 2003 on Canadian English, and McCarthy 2007 and Baković 2011 on opacity, derivations and OT). In the conclusion, I discuss some directions for further research that may help to tease apart future theories of the representation and learning of patterns like Canadian raising.

2. Analysis

The basic distribution of the diphthongs can be analyzed by adapting to HG the standard OT approach to allophony (see McCarthy 2008 for a tutorial introduction). The first ingredient in the analysis is a constraint that prefers the phones with the broader distribution, here [ai] and [au], by penalizing the contextually restricted variants, here [ʌi] and [ʌu]. This constraint, *RAISED, conflicts with a context-specific constraint against the sequence of a low diphthong and tautosyllabic voiceless consonant, *(LOW, VOICELESS). The tableau in (2) shows the situation in

---

when no consonant follows, and generates voicing alternations, rather than raising (see Idsardi 2006 on the productivity of raising).

which *(Low, Voiceless) has its effect, in choosing a raised diphthong before a voiceless consonant.

(2) *(Low, Voiceless) > *Raised + IdHeight

/saik/	*(Low, Voiceless) 4	*Raised 2	IdHeight 1	*H*
☞ [sʌik]		–1	–1	–3
[saik]	–1			–4

The input UR for *psych* [sʌik] is assumed to have low diphthong, from *psychology* with [ai] (this could equally be a richness of the base tableau – see (3) below). Violation counts are shown in the tableaux as negative integers. The correct SR violates the faithfulness constraint penalizing a change in diphthong height, IdHeight, as well as *Raised. Its competitor [saik] violates only *(Low, Voiceless). In HG, the well-formedness, or Harmony of a representation is the weighted sum of its violation scores, shown in the column labeled *H* in the tableau. In an OT-like categorical version of HG, the optimum is the candidate with the highest Harmony. For [sʌik] to beat [saik], the weight of *(Low, Voiceless) must be greater than the summed weights of *Raised and IdHeight – this weighting condition is shown as the caption of the tableau in (2). Weights meeting this condition are shown beneath the constraint names; in section 3 I will discuss one way of finding correct weights using a learning algorithm.

If *Raised has a greater weight than IdHeight, underlying /ʌi/ and /aʌ/ will map to surface [ai] and [au]. This is illustrated in the tableau in (3) for underlying /ʌi/.

(3) *Raised > Id-Height

/ʌi/	*Raised 2	IdHeight 1	*H*
[ʌi]	–1		–2
☞ [ai]		–1	–1

This is a richness of the base tableau, showing that if the grammar is supplied with a raised diphthong that would surface in an inappropriate context, it will map it to the correct low diphthong.

To license the contrast in pre-flap context, we need only add a constraint against low diphthongs in that environment, *(Low, Flap), which in HG can act in a gang effect with IdHeight to counteract *Raise. To show this, I use Prince's (2000) comparative tableau format. The rows in (4) show the differences between the scores of the desired optima (or winners) and their competitors (or losers). The scores of the losers are subtracted from those of the winners, so that winner-preferring constraints display a positive number in the relevant row, while loser-preferring constraints display a negative value. For instance, the first row with Input /saik/ is based on the candidates in (2). The candidate with the raised diphthong, which is the winner here, has no violation of *(Low, Voiceless) while its competitor has one, so the comparative vector has +1. For *Raised and IdHeight the winner has a violation and the loser has none, and the vector shows –1. The second row corresponds to the tableau in (3), and the last two rows correspond to *tidal* and *title*, pronounced as [taiɾl] and [tʌiɾl] in Canadian English.

(4) *Comparative HG tableaux*

Input	W ~ L	*(LOW, VOICELESS)	*RAISED	*(LOW, FLAP)	IDHEIGHT	∑
		4	2	2	1	
/saik/	[sʌik] ~ [saik]	+1	−1		−1	+1
/ʌi/	[ai] ~ [ʌi]		+1		−1	+1
/taiɾl/	[taiɾl] ~ [tʌitl]		+1	−1	+1	+1
/tʌiɾl/	[tʌiɾl] ~ [taiɾl]		−1	+1	+1	+1

As Prince (2000) explains, the comparative format is useful because the OT ranking conditions can be directly read from it. For each row, some constraint that prefers the winner must dominate all constraints that prefer the loser. Here there is no ranking that respects all of these conditions: the last two rows require ID-HEIGHT to dominate *RAISED, as well as *(LOW, FLAP), while the second row requires the contradictory *RAISED >> ID-HEIGHT.

The conditions on a correct HG weighting can also be read from comparative vectors: the sum of the scores in each row, each times the constraint's weight, must be greater than zero (see Potts et al. 2010 for a linear programming method that finds correct weights and detects inconsistency, as well as Pater 2014 for further discussion and references). A set of weights meeting all of the conditions is shown underneath the constraints names in (4), and the final column labeled ∑ shows the weighted sum for each row, which is above zero in each case.

Since all of the non-zero scores assigned by each constraint to each candidate in (4) are +1 and −1, we can also simply say that the sum of the weights of the constraints preferring the winner must be greater than the sum of the weights preferring the loser. For the last two rows in (4), this means that the sum of the weights of IDHEIGHT and *RAISED must be greater than the weight of *(LOW, FLAP), and that the sum of IDHEIGHT and *(LOW, FLAP) must be greater than *RAISED. These conditions do not contradict the need to weight *RAISED above IDHEIGHT, as the second row demands. Because IDHEIGHT can gang up with *(LOW, FLAP) to overcome *RAISED and pick faithful [tʌiɾl] over [taiɾl], it can remain beneath *RAISED, allowing unfaithful [ai] to beat [ʌi] for /ʌi/.

Some refinements remain necessary to handle the full set of data, but insofar as these do not differentiate the current analysis from others, I will only comment on them briefly.[2] First, raising is prosodically conditioned (Paradis 1980): it applies only before tautosyllabic consonants (e.g. *psych* [ʌi] vs. *psy.chology* [ai]).[3] This presumably indicates that the prosodic context must be specified in *(LOW, VOICELESS), requiring the two target segments to be in the same syllable.

---

[2] Idsardi (2006: 26) does offer a piece of data that challenges the current account: he claims that in his own speech there is raising conditioned by underlying /t/ in phrases like "don't lie to me", even when the /t/ of "to" surfaces as a flap. Canadian Raising is usually described as word-bounded, with contrasts such as *lifer* with [ʌi] and *lie for me* with [ai] (Chambers 2011; see *esp.* Bermúdez-Otero 2003), and this fits with my own impressions. More controlled study of varieties like Idsardi's would no doubt be illuminating.

[3] For cases like *Nike* [nʌiki], the tautosyllabicity condition requires that intervocalic post-stress consonants be codas or ambisyllabic. There seems to be some variation in the pronunciation of words like *Cyclops* and *micron* in which a main stress is directly followed by a secondary (Mielke et al. 2003); this may indicate variation in syllabification in this environment.

Second, Bermúdez-Otero (2003) points out that there is also evidence of morphological conditioning (*eye-full* [ai] vs. *Eiffel* [ʌi]), and suggests an analysis in stratal OT. A stratal analysis could be adopted here, but so could another approach, such as writing the morphological or prosodic context into the constraint. Finally, a full account would deal with the paradigm uniformity in the alternations: derived words consistently retain the diphthongal height of their bases. Richness of the base allows for underlying /snait/, which would surface as [snʌit] and [snaiɾɚ], if the latter were derived directly from /snait+ɚ/ with the present constraint set. Presumably, native speakers would reject this novel paradigm as ungrammatical, would fail to learn it, and would regularize it. Hayes (2004) provides an output-output faithfulness account, and Bermúdez-Otero (2003) provides an analysis based on cyclicity. Either approach could be adopted here.

As mentioned in the introduction, this analysis of the distribution of the diphthongs is in a sense intermediate between Joos' (1942) phonemic analysis, and Harris (1951/1960) and Chomsky's (1964) opaque allophonic analyses: [ʌi] and [ai] are contrastive in the pre-flap environment, but nowhere else. Idsardi (2006: 24) objects to Mielke et al.'s (2003) similar postulation of a pre-flap contrast on the basis that it 'marks a return to Joos's view of more than 60 years ago, and is thus subject to all of the criticisms voiced by Chomsky and Harris'. This objection misses the mark, at least for the present analysis. In Joos's phonemic analysis, there is no account of the distribution of the diphthongs. The missed generalizations of that account (and of Mielke et al. 2003; see fn. 1 above) are good reasons to prefer the opaque allophonic analyses offered by Harris and Chomsky and championed by Idsardi (2006). There are, however, no such missed generalizations here: raised vowels are correctly limited to the pre-flap and pre-voiceless consonant environments, and the low diphthongs are correctly banned before voiceless consonants.

To appreciate how and why this analysis differs from the traditional rule-based one, it is important to recognize that the phoneme has no formal status in OT, or in the HG derivative adopted here. Even in the analysis of pure allophony, such as the distribution of flap *vs*. [t] or [d] in North American English, there is no stipulation that only a single phoneme (e.g. /t/ or /d/) may appear in URs. A correct grammar for a given language must map all of the universally possible input URs to just those surface structures that are allowed in that language (= richness of the base) – see (3) above for an example with /ʌi/ in a non-raising environment. The difference between flapping and raising is simply a difference between the weightings of the constraints. IDHEIGHT has sufficient weight to maintain contrast in diphthong height in the pre-flap environment, while the faithfulness constraint(s) relating flap to /t/ and /d/ have insufficient weight to maintain contrast anywhere.[4]

---

[4] A reviewer asks about further richness of the base forms, with an underlying flap word-finally following a raised diphthong (e.g. /fʌiɾ/), and with prevocalic /ʌit/ and /aid/ (e.g. /mʌidɚ/ and /saitɚ/), wondering how the analysis of raising interacts with the analysis of flapping. To answer this question, I added to the constraints in the text ones needed for the analysis of stop/flap allophony, and created tableaux with the three URs above mapping to winners [fʌit], [mʌiɾɚ] and [saiɾɚ]. Each had three competing candidates consisting of all the further combinations of the stop or flap and the two diphthongs. I submitted these tableaux, along with the *eye* tableau of (3) to OT-Help (Staubs et al. 2010). With the minimum weight set to zero, the constraints in the text wound up getting the same weights as in (4) in the text. The new contextual constraint against intervocalic coronal stops was set to 4, and the new context-free constraint against flaps to 2. The faithfulness constraint against stop-flap changes received zero weight. The OT-Help input file can be found at http://people.umass.edu/pater/raising-flap-ot-help.txt.

3. Learning

The preceding analysis assumed a constraint *(LOW, FLAP) that is arbitrary from a phonetic viewpoint, and that is presumably acquired based on a learner's experience with the sound pattern of the language. In this section, I show that the structure of HG, and an associated learning algorithm, allows for a straightforward account of how this constraint, and others, might be induced from the learning data (see also Moreton 2010 and Pizzo 2013 on constraint induction and on-line learning of HG).

In cognitive modeling and machine learning there is a set of widely used closely related learning algorithms called stochastic gradient ascent, the delta rule, and the perceptron update rule. Jäger (2007), Pater (2008), and Boersma and Pater (2014) show that this type of learning can be straightforwardly applied to HG, and in this context closely resembles Boersma's (1998) gradual learning algorithm. The learner is given a correct input-output pair, and uses the current grammar to generate its own optimal output. If the learner's own output fails to match the learning datum, the constraint weights are updated in favor of the correct form. This is done by first subtracting the violations of the correct form, or WINNER from those of the learner's own LOSER, that is, by making a comparative vector of the type used in the analysis in section 2. The values on the vector are scaled by a constant (the learning rate), and are then added to the constraint weights to produce the post-update values. Since constraints favoring the winner have positive values, and constraints favoring the loser negative ones, the update goes in the direction of a set of weights that will make the winner correctly optimal.

This winner-loser comparison can also provide the basis for picking constraints. Let us assume that in updating, the learner also constructs a set of potential constraints by applying a schema to the phonological structures of the winner and the loser. For present purposes, we require only constraints that penalize single segments bearing a single specified feature, and ones that penalize pairs of segments, each with a specified feature; this is a simplified version of a constraint induction schema proposed by Hayes and Wilson (2008).[5] The simplicity of this schema is in part due to the fact that the features I am using in the constraints are themselves abbreviations: 'Low' is a cover feature for that set of features characterizing the lower diphthongs, 'Voiceless' is a cover feature for the set of features characterizing a voiceless consonant, and so on. Given the small set of features I used in the analysis in section 2, the string [ʌik] would project the set of constraints *RAISED, *VOICELESS, and *(RAISED, VOICELESS), while [aik] would project *LOW, *VOICELESS, and *(LOW, VOICELESS). If these two strings formed a Winner-Loser pair, the comparative vector formed over this set of potential output constraints would be as in (5). Note that *VOICELESS favors neither the winner nor loser, since it is violated by both.

---

[5] Hayes and Wilson (2008) show that a somewhat expanded version of this schema for constraints can capture a wide range of phonological patterns, if the constraints also operate over autosegmental and metrical projections. See Kager and Pater 2012 for an argument that the Hayes/Wilson schema must be further expanded with prosodic conditioning (as discussed in section 2, this is also required by the full set of Canadian raising data). See also Hayes and Wilson's argument for implicational constraints, which may well also prove useful in the present approach.

(5) *Comparative vector for potential output constraints from one W ~ L pair*

W ~ L	*LOW	*RAISED	*VOICELESS	*(LOW, VOICELESS)	*(RAISED, VOICELESS)
[ʌik] ~ [aik]	+1	−1		−1	+1

This winner-loser comparison allows the learner to pick only those constraints that favor the winner: here *LOW and *(RAISED, VOICELESS). If the learner had simply projected a set of constraints from the loser, the useless *VOICELESS constraint would also be included.

The constraint set does still have some redundancy, since as we know from the analysis in section 2, the *LOW constraint is not required. From the viewpoint of a theory with violable constraints and a learning algorithm that will order them correctly, this is just a redundancy, and not necessarily a problem. This points to an interesting consequence of the structure of HG and its associated gradual learning algorithm: constraint induction can potentially be done from individual pieces of learning data rather than in terms of calculations over the entire dataset and space of possible constraints, as in Hayes and Wilson (2008). Generalization across the dataset occurs as a function of constraint weighting. In this case, *LOW will be given relatively low weight so that forms like [ai] will be correctly chosen over competing candidates like [ʌi], while *(RAISED, VOICELESS) will be given relatively high weight to pick forms like [sʌik] over [saik]. I will now illustrate this sort of generalization by weighting with a small learning simulation.

For the rime portions of the winner-loser pairs from the analysis in (4), the full set of constraints projected according to schema just described, and which survive the winner-loser comparison also just described and illustrated in (5), are given in (6).

(6) *Induced output constraints*

*LOW, *RAISED
*(LOW, VOICELESS), *(LOW, FLAP), *(RAISED, FLAP)

All of these constraints were included from the beginning of the simulation, since under the current assumptions they would be induced as soon as all four winner-loser pairs were encountered. These constraints, and IDHEIGHT, were given an initial weight of zero, and the learning rate was set to 1. If the update rule produced a value beneath zero, it was adjusted to zero. A typical final set of weights is shown in (7).

(7) *Learned weights*

5	*(LOW, VOICELESS)
3	*RAISED
3	*(LOW, FLAP)
1	IDHEIGHT
0	*LOW, *(RAISED/FLAP)

The learned analysis has the same structure as the one in the previous section. *(LOW, VOICELESS) has a greater weight than *RAISED, which itself outweighs IDHEIGHT: low diphthongs are banned before voiceless consonants, and raised diphthongs are banned elsewhere. The ban on raised diphthongs is subverted before flaps by the gang effect between *(LOW, FLAP) and IDHEIGHT, and the gang effect between *RAISED and IDHEIGHT protects low diphthongs from *(LOW, FLAP), thus producing the pre-flap contrast. The two output constraints that do not play a role in the analysis, but which prefer individual Winners, and were thus created by the induction procedure, both have a weight of zero. This is a simple demonstration that constraints could potentially be projected from individual pieces of data, with generalization across the dataset being taken care of by subsequent applications of the update rule.

Redundant constraints are not guaranteed to receive a zero weight as they did in the simulation above. A good question for further research is whether such redundancies in a fuller constraint set might serve to help explain the direction of language change. There are varieties of American English with Canadian raising in which contrasts are found outside of the pre-flap environment (Vance 1987, Dailey-O'Cain 1997, Fruehwald 2007; Chambers 2011 provides a useful summary). The environments do not seem random – raised vowels have been reported before coronal nasals (e.g. *diner*), semi-syllabic [r] (e.g. *fire*) and voiced stops (e.g. *tiger*). No contrasts have been reported before voiceless consonants, or in the absence of a following consonant. The immediate consonantal environment for the novel contrasts seems to the set of consonants that are featurally close to the flap, and the following vocalic nucleus is often [ɚ], as in many of the Canadian English words with raised vowels without following voiceless consonants. The question is whether contrast in the attested environments is more harmonic than where it does not occur, and whether the development of these contrasts can be modeled with interacting probabilistic learners.[6]

4. Discussion

In the introduction, I justified the use of language-specific, phonetically arbitrary constraints in the analysis of the distribution of Canadian English diphthongs by pointing to languages with phonetically arbitrary phonological patterns. One might well ask why phonological patterns should generally tend to make phonetic sense if phonological constraints (or rules) are not required to be phonetically sensible. One answer, given by Blevins (2004) amongst many others, is that categorical phonological processes arise diachronically from gradient phonetic ones. This process of phonologization is examined in detail for a case of /ai/-raising by Moreton and Thomas (2007), who document and analyze its gradual emergence in Cleveland. They argue that the analysis supports their asymmetric assimilation hypothesis about rising diphthongs, under which the nucleus tends to assimilate to the glide before voiceless consonants, providing the seeds of the Canadian raising pattern, but the glide tends to assimilate to the nucleus before voiced consonants, which can yield a Southern American English pattern in which [ai] is found only before voiceless consonants. They also argue that their data run counter to the predictions of other accounts of the diachronic source of Canadian raising, which claim that it comes from the differential application of the great vowel shift before voiced and voiceless consonants, and/or that it results from the relative shortness of vowels before voiceless consonants.

---

[6] As a reviewer notes, the development of contrast is also potentially related to a shortcut that I took in the learning simulations. The URs for learning were as shown in (4), including the unfaithful /saik/ for [sʌik] and /ʌi/ for [ai]. This is a simple, albeit unrealistic, way of getting the desired result in which the underlying specification of the height of the diphthong in these environments does not matter for the surface outcome. If we run the simulation as described in the text with faithful underlying forms, IdHeight winds up getting the greatest weight, leading to predicted contrast in all environments (see relatedly Hayes 2004). This does not seem appropriate for Canadian English, since there is no evidence of contrast outside of the pre-flap environment. A full account would include mechanisms for the learning of URs, and of sufficiently restrictive grammars – see Pater *et al.* 2012 and Jesney and Tessier 2011 respectively for recent proposals in HG and references to other work.

The main point of the present paper is that if we allow language-specific phonetically arbitrary constraints or rules, then patterns that usually are claimed to require intermediate derivational stages and abstract underlying representations, of which Canadian raising and flapping is the canonical case, can at least sometimes be analyzed without these formal mechanisms. How many of them can be, whether they should be, and whether these analyses should use HG, rather than OT, a rule-based framework, or some other approach, are all questions for further research.

In pursuing a research program that aims to explain phonological typology in terms of the interaction of phonetics, phonology, and language use and learning over time, there is a range of considerations that might lead to picking one phonological framework over another. Chief amongst these is whether it has an associated learning algorithm that allows for the explicit modeling of phonology's place in this system. HG's gradual learning algorithm seems well suited for this task. Another consideration is whether it can help to explain skews in typology that cannot be explained by phonetics alone. One likely set of cases involves tendencies toward systemic simplicity, like feature economy (Clements 2003). Pater and Moreton (2012) and Pater and Staubs (2013) provide some initial results showing that skews toward systemic simplicity emerge from iterated learning using the basic grammar and learning setup adopted here.

Given a set of descriptively adequate theories of the grammatical representation of some set of phonological patterns, for example those that are derivationally characterized as counterbleeding opacity, one can ask how successful they are in explaining learning. Assuming that each of the grammatical theories can be paired with a suitable learning algorithm, this issue can be studied in at least two ways. One is to measure success in reaching correct final states. The present analysis avoids two hidden structure problems that may well hinder success in this regard: the hidden intermediate level of derivation, and the abstract /t-d/ contrast[7] and underlying /ai/ for [ʌi] (though see Bermúdez-Otero 2003). The cost of avoiding these hidden structures might be the complexity of the required constraints: other analyses do not posit constraints penalizing [ai] in pre-flap position. HG seems helpful in minimizing this cost. An analysis in standard OT without intermediate levels would require something like a version of IDHEIGHT that is specific to the pre-flap environment, whereas the HG analysis requires only the simpler general faithfulness constraint, acting in a gang effect with *(LOW, FLAP) – see relatedly Jesney 2011 on positional faithfulness and ranking *vs.* weighting.

Theories of grammar and learning can also be compared for the predictions that they make for the course of acquisition. This seems like a particularly promising way of teasing apart theories of the phenomena termed opacity, with the increasingly popular methodology of artificial language learning providing a means of testing predictions that would likely be impossible to study in the 'wild' (see Moreton & Pater 2012 for a review). Moreton and colleagues (2013) show that the predictions of a weighted constraint model for phonotactic learning are rather strikingly confirmed using this approach, so once again, there is reason to see promise for future HG models.

---

[7] Since there is variation in flapping, there are likely surface-observed variants in this particular case, but this is probably not true of all cases of abstract conditioning segments in counterbleeding opacity.

REFERENCES

ANDERSON, STEPHEN. 1981. Why phonology isn't "Natural". *Linguistic Inquiry* 12. 493-539.

BACH, EMMON AND R. T. HARMS. 1972. How do languages get crazy rules? *Linguistic change and generative theory*, ed. by Robert Stockwell and Ronald Macaulay, 1-21. Bloomington: Indiana University Press.

BAKOVIĆ, ERIC. 2011. Opacity and ordering. *The Handbook of Phonological Theory* (2nd ed), ed. by John Goldsmith, Jason Riggle, and Alan Yu, 40-67. Malden, MA: Wiley-Blackwell.

BERMÚDEZ-OTERO, RICARDO. 2003. The acquisition of phonological opacity. I*Variation within Optimality Theory*. ed. by Jennifer Spenader, Anders Eriksson and Östen Dahl. Department of Linguistics, Stockholm University. 27-38.

BLEVINS, JULIETTE. 2004. *Evolutionary Phonology*. Cambridge University Press.

BOERSMA, PAUL. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Ph.D. dissertation, University of Amsterdam.

BOERSMA, PAUL, AND JOE PATER. 2007. Constructing constraints from language data: The case of Canadian English diphthongs. Handout, NELS 38, Ottawa. Online: http://people.umass.edu/pater/boersma-pater-nels.pdf.

BOERSMA, PAUL AND JOE PATER. 2014. Convergence properties of a gradual learning algorithm for Harmonic Grammar. *Harmonic Grammar and Harmonic Serialism*, ed. by John McCarthy and Joe Pater. London: Equinox Press, to appear.

CHAMBERS, J.K. 1973. Canadian raising. *Canadian Journal of Linguistics* 18. 18-35.

CHAMBERS, J.K. 2011. Learning to love opacity: Progress of /ai/-Raising. Paper presented at the Montreal-Ottawa-Toronto Phonology Workshop, McGill University.

CLEMENTS, G. N. 2003. Feature economy in sound systems. *Phonology* 20. 287-333.

DAILEY-O'CAIN, JENNIFER. 1997. Canadian raising in a midwestern U.S. city. *Language Variation and Change* 9. 107-20.

FRUEHWALD, JOSEF T. 2007. The Spread of raising: opacity, lexicalization and diffusion. *College Undergraduate Research Electronic Journal*. Online: http://repository.upenn.edu/curej/73.

HANNSON. GUNNAR. 2007. On the evolution of consonant harmony: the case of secondary articulation agreement. *Phonology* 24. 77-120.

HARRIS, ZELLIG. 1951/1960. *Structural linguistics*. Chicago: University of Chicago Press.

HAYES, BRUCE. 2004. Phonological acquisition in Optimality Theory: The early stages. *Constraints in Phonological Acquisition*, ed. by René Kager, Joe Pater and Wim Zonneveld, 158-203, Cambridge University Press.

HAYES, BRUCE, KIE ZURAW, PÉTER SIPTÁR, AND ZSUZSA LONDE, 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85. 822-863.

HAYES, BRUCE AND COLIN WILSON. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379-440.

IDSARDI, WILLIAM J. 2006. Canadian raising, opacity, and rephonemization. *Canadian Journal of Linguistics* 51. 21-28.

JÄGER, GERHARD. 2007. Maximum entropy models and Stochastic Optimality Theory. *Architectures, rules, and preferences: A festschrift for Joan Bresnan*. ed. by Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson, and Annie Zaenen, 467-479. Stanford, CA: CSLI.

JESNEY, KAREN. 2011. *Cumulative Constraint Interaction in Phonological Acquisition and Typology.* Ph.D. dissertation, University of Massachusetts Amherst.

JESNEY, KAREN, AND ANNE-MICHELLE TESSIER. 2011. Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language and Linguistic Theory* 29. 251-290.

JOOS, MARTIN. 1942. A phonological dilemma in Canadian English. *Language* 18, 141-144.

KAGER, RENÉ AND JOE PATER. 2012. Phonotactics as phonology: Knowledge of a complex restriction in Dutch. *Phonology* 28, 197-234.

MCCARTHY, JOHN J. 2007. *Hidden Generalizations: Phonological Opacity in Optimality Theory*. London: Equinox.

MCCARTHY, JOHN J. 2008. *Doing Optimality Theory*. Malden, MA: Wiley-Blackwell.

MIELKE, JEFF, MIKE ARMSTRONG AND ELIZABETH HUME. 2003. Looking through opacity. *Theoretical Linguistics* 29. 123-139.

MORETON, ELLIOTT. 2010. Constraint induction and simplicity bias in phonotactic learning. Paper presented at the Workshop on Grammar Induction, Cornell University. On-line: http://www.unc.edu/~moreton/Papers/MoretonCornellWkshp2010.pdf.

MORETON, ELLIOTT AND JOE PATER. 2012. Structure and substance in artificial-phonology learning. Part 1: Structure, Part II: Substance. *Language and Linguistics Compass* 6 (11): 686–701 and 702–718.

MORETON, ELLIOTT, JOE PATER AND KATYA PERTSOVA. 2013. Phonological concept learning. Ms, University of North Carolina and University of Massachusetts Amherst.

MORETON, ELLIOTT, AND ERIK R. THOMAS. 2007. Origins of Canadian raising in voiceless-coda effects: a case study in phonologization. *Laboratory Phonology 9*, ed. by Jennifer S. Cole and José Ignacio Hualde, 37-64. Berlin: Mouton.

PARADIS, CAROLE. 1980. La règle de Canadian raising et l'analyse en structure syllabique. *Canadian Journal of Linguistics* 25. 35–45.

PATER, JOE. 2008. Gradual learning and convergence. *Linguistic Inquiry* 39. 334-345.

PATER, JOE. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33: 999-1035.

PATER, JOE. 2014. Universal Grammar with Weighted Constraints. *Harmonic Grammar and Harmonic Serialism*, ed. by John McCarthy and Joe Pater. London: Equinox Press, to appear.

PATER, JOE AND ELLIOTT MORETON. 2012. Structurally biased phonology: Complexity in learning and typology. *EFL Journal* (The Journal of the English and Foreign Languages University, Hyderabad), 1-44.

PATER, JOE AND ROBERT STAUBS. 2013. Feature economy and iterated grammar learning. Paper presented at the Manchester Phonology Meeting. Online: http://people.umass.edu/pater/pater-staubs-mfm-2013.pdf].

PATER, JOE, ROBERT STAUBS, KAREN JESNEY AND BRIAN SMITH. 2012. Learning probabilities over underlying representations. *Proceedings of the Twelfth Meeting of the ACL-SIGMORPHON: Computational Research in Phonetics, Phonology, and Morphology*, 62-71.

PIZZO, PRESLEY. 2013. Online Constraint Induction for Vowel Harmony. Ms., University of Massachusetts Amherst.

POTTS, CHRISTOPHER, JOE PATER, KAREN JESNEY, RAJESH BHATT AND MICHAEL BECKER. 2010. Harmonic Grammar with Linear Programming: From linear systems to linguistic typology. *Phonology* 27, 77-117.

PRINCE, ALAN. 2000. Comparative tableaux. Ms, Rutgers University. Online: http://roa.rutgers.edu/files/376-0100/roa-376-prince-1.pdf

SCHEER, TOBIAS. 2014. Crazy rules, regularity and naturalness. *The Handbook of Historical Phonology*. ed. by Joseph Salmons and Patrick Honeybone. Oxford University Press, to appear.

SMOLENSKY, PAUL AND GÉRALDINE LEGENDRE. 2006. *The harmonic mind: From neural computation to optimality-theoretic grammar*. MIT Press.

STAUBS, ROBERT, MICHAEL BECKER, CHRISTOPHER POTTS, PATRICK PRATT, JOHN J. MCCARTHY AND JOE PATER. 2010. OT-Help 2.0. Software package. Amherst, MA: University of Massachusetts Amherst. Online: http://people.umass.edu/othelp/

VANCE, TIMOTHY. 1987. ‘Canadian raising’ in some dialects of the northern United States. *American Speech* 6, 195-210.