NELS 38, University of Ottawa, Oct. 27th, 2007

Constructing constraints from language data: The case of Canadian English diphthongs[1]

Paul Boersma and Joe Pater

University of Amsterdam and University of Massachusetts, Amherst

1. Overview

(1) *§ 2 Proposal*
Learning of structural constraints from schema applied to observed forms

§ 3 Illustration
Learning semi-predictable distribution of diphthong pairs [ai] / [ʌi] and [au] / [ʌu] in Canadian English, including natural and unnatural constraints

§ 4 Proposal, Illustration
How natural phonological constraints emerge from phonetic biases in a bidirectional model of phonology and phonetics: phonologization of gradient raising

2. Learning of structural constraints from observed forms

2.1 Background assumptions

Optimality Theory (OT) translated to Harmonic Grammar (HG) (original translation due to Prince and Smolensky 1993/2004: 236; HG and probably the more general notion of numerically weighted linguistic constraints is due to Legendre, Miyata and Smolensky 1990; original application of HG to phonology is Goldsmith 1990/1993; for further references and discussion see Smolensky and Legendre 2006 and Pater, Bhatt and Potts 2007)

(2)

	2	1	*H*
bat	NoCoda	Max	
bat		1	1
☞ ba	1		2

The HG tableau conventions assumed here:

(3) i. The cell above each constraint's name is that constraint's weight

ii. Constraints assign positive scores for satisfaction

iii. The rightmost cell in each candidate's row is that candidate's *Harmony* (sum of weighted constraint scores)

---

[1] Special thanks to Michael Becker for indispensible technical help (AKA OT-Help), and to Karen Jesney with help on native speaker judgments. Thanks also to them and Adam Albright, Ricardo Bermúdez-Otero, Bruce Hayes, Bill Idsardi, Elliot Moreton and Matt Wolf for very useful discussion.

iv. The optimal candidate is the one with the highest Harmony

On-line error-driven OT/HG learner (Tesar 1995 *et seq.*; Boersma 1998 *et seq.*; Jäger 2003; Pater 2007)

(4) i. Correct *input-output* pair (i_1, o_{11}) presented to learner

ii. Learner computes optimal output (o_{12}) given i_1, and its current ranking/weighting of the constraints

iii. If o_{11} does not match o_{12} (an 'error'), learning is triggered

For example, if the correct pair (bat, bat) were presented to a learner with the weighting in (2) above, the indicated optimal pairing (bat, ba) would be an error.

HG update rule (Jäger 2003; Soderstrom, Mathis & Smolensky 2006; Boersma & Weenink 2007; Pater 2007):

(5) For each constraint, the updated value is calculated by subtracting the incorrect mapping's score from that of the correct mapping, multiplying this value by a constant *n*, and adding the result to that constraint's weight

Violation scores of the correct mapping (C) and error (E) in our example, and their difference:

(6)

	NoCoda	Max
C (bat, bat)		1
E (bat, ba)	1	
C - *E*	-1	1

If *n*=1, the result of adding the result of *C* - *E* in (6) to the weights in (2) is shown in (7). As also shown in (7), the grammar will successfully parse the next coda in the learning data:

(7)

	2	1	*H*
bat	Max	NoCoda	
☞ bat	1		2
ba		1	1

Convergence proofs and testing of this learning procedure for HG and HG-like theories of grammar, including ones that deal with variation (unlike standard OT):

(8) Fischer (2005), Boersma and Pater (2007)

To achieve a sufficiently restrictive end-state grammar (Smolensky 1996, Hayes 2004, Prince and Tesar 2004) we assume that the structural constraints (like NoCoda), begin with an initial weight greater than the faithfulness constraints (like Max). Comparison of HG with OT in this regard, including a proposal that the weightings of structural constraints changes faster than those of faithfulness constraints:

(9) Jesney and Tessier (today, 2007), Magri (in prep.)

Other applications of HG linear models and HG-like log-linear models to phonology and phonological learning (amongst others, we're sure):

(10) Goldwater and Johnson (2003), Wilson (2006), Hayes and Wilson (2007), Goldrick and Daland (2006), Coetzee and Pater (2007), Albright, Michaels and Magri (2007)

2.2 The proposal

An issue:

(11) How can an *on-line* learner acquire constraints that generalize across observed forms?

Our answer:

(12) i. For each observed form, the learner creates constraints encoding the relationships between structural elements

ii. 'Generalization' comes from the subsequent re-weighting of constraints as more forms are encountered

Our general conception of constraint construction (see Burzio 2002 on implicational constraints):

(13) Given structural element e_1 and structural element e_2 that co-occur in an observed form, create a constraint:

$e_1 \rightarrow e_2$ Assign a reward of 1 to each instance of e_1 that co-occurs with e_2.

This general idea could be implemented in many ways. The best approach can only be determined through large-scale implementation and testing, which we haven't yet undertaken.

(14) a. Assignment of negative penalties for violation (as in standard OT)?[2]

b. Constraints that refer to sets of elements: what size are the sets?

c. Constraints that refer to elements in particular domains/structural positions: what are the domains/windows within which the constraints are created?

We have been inspired by Hayes and Wilson's (2007) approach to phonotactic learning, which provides concrete proposals about some of the above. The main differences between our approaches (critical comparison would be premature):

---

[2] We state our constraints positively for simplicity. We are aware of the infinite goodness problem. Whether this will be a problem for our system will depend on whether we adopt an OT-style GEN (cf. Soderstrom *et al.* 2006). If we do, the problem can be avoided by simply translating all of our constraints into penalty assigners.

(15) a. We use a grammar of a form familiar from OT (a MaxEnt version would estimate conditional probabilities, as in Goldwater and Johnson 2003, Jäger 2003, Wilson 2006), with an on-line learner

b. Hayes and Wilson create all logically possible constraints fitting their schema, independent of whether they are satisfied in any observed form or not, and then winnow down that set (see relatedly Boersma, Escudero and Hayes 2003)

3. Illustration: Distribution of Canadian English diphthongs

3.1 Generalization through learning and re-weighting constraints

Generalization about Canadian English diphthongs (*to be refined!*):

(16) 'Raised' diphthongs [ʌi], [ʌu] before voiceless consonants, [ai], [au] elsewhere

Given a single observed form [ʌis] 'ice', a learner could create the following constraints, amongst many others (we use 'D' as an abbreviation for the relevant 'root node' characteristics of the diphthongs, and 'C' for consonant):

(17)	*Observed structure*	*Constructed constraint*
a.	D [-low]	RAISED DIPHTHONG: A diphthong must be [-low] (Assign a reward of 1 to each diphthong that is [-low])
b.	D C [-low][-vce]	RAISED/VOICELESS: A diphthong preceding a voiceless consonant must be raised (Assign a reward of 1 to each diphthong that is [-low] that precedes a voiceless consonant)

The constraint that encodes the correct generalization is of course (17b). The learner will discover this after encountering further data, like [ai] 'eye' and [saiz] 'size', which yield the constraints in (18):

(18)	*Observed structure*	*Constructed constraint*
a.	D [low]	LOW DIPHTHONG: A diphthong must be [low]
b.	D C [low][+vce]	LOW/VOICED: A diphthong preceding a voiced consonant must be [low]

We'll make the simplifying assumption that these three structural constraints interact with a single faithfulness constraint:

(19) *Faithfulness constraint*

IDENT-[LOW] — Assign a reward of 1 to an input vowel whose output correspondent has the same specification for Low

We give the structural constraints an initial weighting value of 10, and the Faithfulness constraint an initial value of 0. Learning data (supplied in equal proportions):

(20) (ʌis, ʌis) (ai, ai) (saiz, saiz)

The constraint values found by the Praat HG learning implementation (Boersma and Weenink 2007) are shown in the following tableaux.

(21)

		11	10	10	9	1	
Input	Output	LOW DIPHTHONG	RAISED / VOICELESS	RAISED / VOICED	RAISED DIPHTHONG	IDENT-LOW	*H*
ʌis	☞ ʌis		1		1	1	20
	ais	1					11
ai	☞ ai	1				1	12
	ʌi				1		9
saiz	☞ saiz	1				1	12
	sʌjz			1	1		19

Idsardi (2006) provides the following clever example incontrovertibly demonstrating productivity (and also that richness of the base must yield raising, rather than devoicing of voiced consonants following raised vowels, as in Mielke, Armstrong and Hume 2003):

(22) '*i*' [ai] '*y*' [wai] *versus* '*i*-th' [ʌiθ] '*y*-th' [wʌiθ]

We derive the raising pattern: not only do these values pick the correct forms, they also produce the correct results in a richness of the base test,[3] where the diphthongs are contextually misplaced in the input.

[3] We do not wish to imply that all things done under the rubric of 'richness of the base' are necessarily psychologically real. In the absence of alternations like those in (23), this can be interpreted as shorthand for what a speaker does in performing a phonological grammaticality judgment task: mapping of an observed form to an identical underlying form, and then creating an input-output map to see if it matches (as in Coetzee 2004; this is similar to the procedure in error-driven learning described above). Such a task may or may not be possible in practice, depending on whether speakers are able to robustly perceive the phonologically ill-formed structure.

(23)

		11	10	10	9	1	
Input	Output	Low Diphthong	Raised / Voiceless	Low / Voiced	Raised Diphthong	Ident-Low	*H*
ais	☞ ʌis		1		1		19
	ais	1				1	12
ʌi	☞ ai	1					11
	ʌi				1	1	10
sʌiz	☞ saiz	1		1			21
	sʌiz				1	1	10

A simple point here, which is common to this approach and the standard OT account of learning, with a large universal constraint set (Tesar 1995 *et seq*.):

(24) Generalization comes from constraint ranking/weighting

A more novel point (though see precedents discussed above):

(25) Phonological constraints can be directly induced from observed forms

3.2 An application of constraint learning: a 'crazy contrast'

Some arguments that constraints must be learned:

(26) i. Claims about innateness require better motivation than typological observations about things like attestation of post-nasal voicing and absence of pre-nasal voicing (see Pater 1995/1999/2004 on the typology, and Hayes and Stivers 1995 and Hayes 1999 on the phonetics and a learning proposal; note that the claim in Pater 1999 is that the constraint is universal, not innate - perhaps better would be 'universally available'. Note too that Prince and Smolensky 1993/2004 seem never to use the word 'innate' (though they do speak of UG); see Tesar and Smolensky 2000 for relevant discussion)

ii. Morpheme-specific constraints (first posited in OT for Lardil and Tagalog in Prince and Smolensky 1993/2004) are necessarily language-specific (see Pater to appear for a learning proposal)

iii. If phonological categories are learned, then the constraints that refer to them must be learned (Boersma 1998; see Maye 2000 and Mielke 2004 for related work)

iv. Needed to analyze 'crazy rules' (Bach and Harms 1972 *et seq.*)

Here we demonstrate another consequence of learned constraints:

(27) Analysis, including learning, of 'crazy contrasts'

The distribution of Canadian English diphthongs includes an instance of a crazy contrast (see Joos 1942, Vance 1982, Mielke, Armstrong and Hume 2003, and Hayes 2004 for variants of contrast-based analyses, see

Chomsky 1957, Halle 1962, Chomsky 1964, Chomsky and Halle 1968, Chambers 1973/1975 and Idsardi 2006 (source of references) for an alternative approach, and Bermúdez-Otero 2003 for yet another):

(28) *Contrast before flaps*

[tʰʌiɾl̩]	title
[bɹaiɾl̩]	bridle
[mʌiɾɹ̩]	mitre
[saiɾɹ̩]	cider

This is the only segmental[4] environment for the Canadian[5] English contrast. This contrast is 'crazy' in that there seems to be no phonetic reason that the flap should license it.

The challenge (Bermudez-Otero 2003):

(29) To show how a learner can arrive at a grammar that allows the pre-flap contrast, but still productively chooses the correct form of the diphthongs elsewhere

Some more evidence of productivity (Note that some of these could be lexicalized, and that they haven't all been well checked. Note too that 'psychology' and other similar examples will require an expanded constraint set; we have experimented with a syllable-based analysis, which seems to work fine):

(30) a. 'south' [ʌwθ] *versus* 'southie' [awð]
b. 'house' [ʌws] *versus* 'house' [awz] (also alternate plural form 'houses' [hawzəz])
c. 'life' [ʌjf] *versus* 'lives' [ajv]
d. 'psych' [ʌjk] *versus* 'psychology' [aj.k]
e. 'scyth' [ʌjθ] *versus* 'scythes' [aiðz]

Presenting the words in (28) to the learner yields two more constraints, amongst many others:

(31)

Observed structure	*Constructed constraint*
D C [-low][flap]	RAISED / FLAP A raised diphthong is followed by a flap

---

[4] In the second author's speech, it appears to be the case that a diphthong in a primary stressed syllable followed by a secondary stressed syllable is invariably unraised (e.g. *icon, cyclops, micron, eyeful*). However, there are reports of contrast and morphological influences in this environment (see Bermudez-Otero 2003 and Mielke et al. 2003 for discussion and references), and informal observation seems that to indicate that there may even be intra-speaker variation conditioned by speech rate. Clearly, Canadian raising provides a continuing source of opportunities for both empirical and theoretical research.

[5] See Vance (1987) for evidence of contrast in upstate New York. We have not come across similar data in Canadian speech.

D C	LOW / FLAP
[+low][flap]	A low diphthong is followed by a flap

A difference between HG and OT is that we can get the contrast as a *gang effect* between these constraints and IDENT-[LOW], as exemplified in the following tableaux:

(32)

	2	2	1	*H*
tʰʌjɾl̩	RAISED / FLAP	LOW / FLAP	IDENT-LOW	
☞ tʰʌjɾl̩	1		1	3
tʰajɾl̩		1		2

(33)

	2	2	1	*H*
bɹaiɾl̩	RAISED / FLAP	LOW / FLAP	IDENT-LOW	
☞ bɹaiɾl̩		1	1	3
bɹʌiɾl̩	1			2

Because IDENT-LOW can maintain a low position in the constraint hierarchy and produce contrast, in our slightly larger learning simulation, we still get the emergent alternations.

The constraint set and learned values (RAISED / VOICED comes from [tʰʌjɾl̩]):

(34)

LOW DIPHTHONG	12
RAISED / VOICELESS	11
RAISED / FLAP	11
RAISED / VOICED	11
LOW / FLAP	9
RAISED DIPHTHONG	8
LOW / VOICED	8
IDENT-VOICE	3
IDENT-LOW	2
IDENT-FLAP	0

The observed mappings (both in the learning data, and final state output):

(35) (bɹaiɾl̩, bɹaiɾl̩)
(tʰʌiɾl̩, tʰʌiɾl̩)
(saiz, saiz)
(ai, ai)
(ʌis, ʌis)

Generalization to unseen forms:

(36) (aid, aid)
(ʌid, aid)
(ait, ʌit)
(ʌit, ʌit)
(ʌiz, aiz)
(ai, ai)
(ais, ʌis)

4. The emergence of universal and language-specific constraints

So we maintain that structural constraints express language-specific structural facts about representations. The big question then becomes: how come so many languages have the same kind of structural constraints, and the same kinds of rankings/weightings? Here we give a schematic answer.

Since pre-voiceless raising of low vowels seems to have emerged independently in various places (for an overview, see Moreton & Thomas 2004), Canadian raising is a good case for answering the universality question. We minimally need the following model of connections within and between the phonology and the phonetics (Boersma 1998, 2007; articulatory form and auditory form collapsed into one phonetic form for simplicity):

(37)

<morphemes>
lexical constraints
|underlying form|
faithfulness constraints
/surface form/ structural constraints
cue constraints
[phonetic form] articulatory constraints

4.1 Generation 1 optimizes her comprehension

Suppose that the first generation of learners is confronted with a ‘perfect’ language environment in which the vowels in <ice> and <size> are pronounced with the exact same quality [ɐi], despite their different durations:

(38) *Language environment for a learner of generation 1*
<ice> [ɐ̆is]
<size> [sɐ̄iz]

The learner’s task as a listener is to make sense of these given combinations of sound and meaning. She will have to construct the intermediate representations (underlying and surface form). The learner’s first action will be to classify the sounds into appropriate categories. On the basis of [ɐ̆i] and [ɐ̄i] she will have no reason to posit any different category of vowel *quality* than /ɐi/. The learner’s sound input will be highly variable, i.e. for the vowel /ɐi/ it will contain realizations like [ɐ̞i], [ɐi], and [ɐ̝i]. On the basis of the language data, then, the learner will **create new cue constraints** for all possible combinations of sound and category (Boersma, Escudero & Hayes 2003:1015):

(39) *Cue constraints for vowel quality created by a learner of generation 1*
/ɐi/[ɐ̞i]
/ɐi/[ɐi]
/ɐi/[ɐ̝i]

It has been shown in computer simulations that a Stochastic OT or Stochastic HG learner who optimizes her comprehension (i.e. her capability of disambiguating vowels) will end up ranking her cue constraints in such a way that they prefer more peripheral auditory values over more central auditory values (Boersma 2006a, Boersma & Hamann 2007). For the case at hand, where peripheral means 'low', the cue constraints will end up being weighted higher for lowered realizations than for raised realizations:

(40) *Comprehension-optimizing weighting of cue constraints for a learner of generation 1*
/ɐi/[ɐ̞i] >> /ɐi/[ɐi] >> /ɐi/[ɐ̝i]

The learner will also create cue constraints for the relation between vowel duration and voicing. The duration of the vowel before voiceless consonants is assumed to have been phoneticized in an English-specific way (i.e. any automatic articulatory or acoustic motivation has been lost), and is therefore regulated by cue constraints that are weighted in an English-specific way (Boersma 2006c):

(41) *Cue constraints for vowel duration created by a learner of generation 1*
/V/[V̆−vce] >> /V/[V−vce] >> /V/[V̄−vce]
/V/[V̄+vce] >> /V/[V+vce] >> /V/[V̆+vce]

The learner's second action is to link the perceived vowel /ɐi/ to something in the lexicon, presumably preferably the underlying form |ɐi|. The learner can do this by **creating new faithfulness constraints** as soon as she has created the relevant categories (Boersma 1998:278), which are here only /ɐi/:

(42) *Faithfulness constraint created by a learner of generation 1*
|ɐi|/ɐi/

Finally, stored underlying forms have to be linked to morphemes (lemmas) and to meaning. As soon as a morpheme cooccurs with an underlying form, the learner can **create a new lexical constraint** (Boersma 2001, Escudero 2005:214–236, Apoussidou 2006). For instance, one lexical constraint connects the morpheme <ice> to the underlying form |ɐis|. The speaker has not encountered any other UFs for <ice>, nor any other morphemes for |ɐis|; hence she needs only a single positive constraint for each connection:

(43) *Lexical constraints created by a learner of generation 1*
<ice>|ɐis|
<size>|sɐiz|

If these constraints are available, and the learner hears the sound [ɐ̆is], she will be able to access the correct morpheme, as shown by the following parallel comprehension tableau:

(44) *Parallel comprehension in generation 1*

	10	8	7	7	5	3	2	*H*
[ɐ̆is]	<ice> |ɐjs|	|ɐj| /ɐj/	/V/ [V̆−vce]	/V/ [V̄+vce]	/ɐi/ [ɐ̞i]	/ɐi/ [ɐi]	/ɐi/ [ɐ̝i]	
☞ <ice> |ɐis| /ɐis/	+1	+1	+1			+1		+28

Given the positive constraints mentioned in this section, and restricting GEN in such a way that every connection between adjacent representations in (41) has to positively satisfy at least one constraint, the tableau can contain only one candidate, and it is a quite good one. The only thing that would be even more harmonic (*H* = +30) is the comprehension of the enhanced (motherese?) utterance [ɐ̞̆is]...

4.2 Generation 1 introduces a phonetic contrast in production

The generation-1 learner has become successful as a listener. But she will also become a speaker herself, and during a phase of sensorimotor acquisition she will learn that the sounds [ɐ̞], [ɐ], and [ɐ̝] can be implemented by the articulations [ɐ̞], [ɐ], and [ɐ̝], respectively. On the basis of vocal play, then, she will **create new articulatory constraints** (Boersma 1998:280), separately for long and short realizations:

(45) *Articulatory constraints created by a learner of generation 1*
*[ɐ̞̆] *[ɐ̞̄]
*[ɐ̆] *[ɐ̄]
*[ɐ̝̆] *[ɐ̝̄]

We note that in speaking she will have to take into account that shortened vowels (such as typically occur in English before voiceless consonants) are more difficult to pronounce with a fully open jaw. This leads to a fixed weighting of articulatory constraints that prefer raised realizations over lowered realizations (if they are short) and long realizations over short realizations (if they are lowered):

(46) *Effort-based weighting of articulatory constraints for speakers of generation 1*
*[ɐ̞̆] >> *[ɐ̆] >> *[ɐ̝̆]
*[ɐ̞̆] >> *[ɐ̞̄]

We next have to assume **bidirectional use of constraints**, i.e. in production the learner will use the same weightings that have optimized her comprehension earlier (Boersma 2007, 2006a; Boersma & Hamann 2007). The cue constraints then come to interact with the articulatory constraints. With a certain intertwining of these two constraint families, the <ice> morpheme will be pronounced with a raised diphthong:

(47) *Parallel production in generation 1*

<table>
<tr><td></td><td>10</td><td>8</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td><td>H</td></tr>
<tr><td><ice></td><td><ice>
|ɐis|</td><td>|ɐi|
/ɐi/</td><td>/V/
[V̆−vce]</td><td>*[ɐ̝̆]</td><td>/ɐi/
[ɐ̝i]</td><td>*[ɐ̆]</td><td>/ɐi/
[ɐi]</td><td>/ɐi/
[ɐ̝i]</td><td>*[ɐ̝̆]</td><td></td></tr>
<tr><td>☞ |ɐis| /ɐis/ [ɐ̝̆is]</td><td>+1</td><td>+1</td><td>+1</td><td></td><td></td><td></td><td></td><td>+1</td><td>−1</td><td>+26</td></tr>
<tr><td>|ɐis| /ɐis/ [ɐ̆is]</td><td>+1</td><td>+1</td><td>+1</td><td></td><td></td><td>−1</td><td>+1</td><td></td><td></td><td>+24</td></tr>
<tr><td>|ɐis| /ɐis/ [ɐ̝̆is]</td><td>+1</td><td>+1</td><td>+1</td><td>−1</td><td>+1</td><td></td><td></td><td></td><td></td><td>+24</td></tr>
<tr><td>|ɐis| /ɐis/ [ɐ̝̄is]</td><td>+1</td><td>+1</td><td></td><td></td><td>+1</td><td></td><td></td><td></td><td></td><td>+23</td></tr>
</table>

The voiced case, unhindered by any high-weighted articulatory constraints, ends up with the low vowel that is preferred by the cue constraints:

(48) *Parallel production in generation 1*

<table>
<tr><td></td><td>10</td><td>8</td><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td><td>H</td></tr>
<tr><td><size></td><td><size>
|sɐjz|</td><td>|ɐj|
/ɐj/</td><td>/V/
[V̄+vce]</td><td>*[ɐ̝̆]</td><td>/ɐi/
[ɐ̝i]</td><td>*[ɐ̆]</td><td>/ɐi/
[ɐi]</td><td>/ɐi/
[ɐ̝i]</td><td>*[ɐ̝̆]</td><td></td></tr>
<tr><td>|sɐiz| /sɐiz/ [sɐ̝̄iz]</td><td>+1</td><td>+1</td><td>+1</td><td></td><td></td><td></td><td></td><td>+1</td><td></td><td>+27</td></tr>
<tr><td>|sɐiz| /sɐiz/ [sɐ̄iz]</td><td>+1</td><td>+1</td><td>+1</td><td></td><td></td><td></td><td>+1</td><td></td><td></td><td>+28</td></tr>
<tr><td>☞ |sɐiz| /sɐiz/ [sɐ̝̄iz]</td><td>+1</td><td>+1</td><td>+1</td><td></td><td>+1</td><td></td><td></td><td></td><td></td><td>+30</td></tr>
<tr><td>|sɐiz| /sɐiz/ [sɐ̝̆iz]</td><td>+1</td><td>+1</td><td></td><td></td><td>+1</td><td></td><td></td><td></td><td>−1</td><td>+22</td></tr>
</table>

Within a single generation, then, the language has introduced a small phonetic bias on the basis of differential articulatory effort: low vowels get raised before voiceless consonants. (Comment: the present explanation involved duration; other explanations in terms of articulatory effort, biomechanics, physiological-acoustic relations, or acoustic transmission would give similar results.)

4.3 Generation 2 has a full phonetic contrast (i.e. in comprehension and production)

The children of the learner of generation 1 are confronted with a slightly different language environment than their parents:

(49) *Language environment for a learner of generation 2*
<ice> [ɐ̝̆is]
<size> [sɐ̝̄iz]

If the difference between the two realizations is not large enough to create two separate vowel quality categories, the learner will create the same category /ɐi/ and therefore the same faithfulness and lexical constraints as her parents did. She will also create the same cue constraints for vowel duration. But there are new kinds of variation: the <ice> vowel will vary between [ɐ̆i], [ɐ̝̆i], and [ʌ̆i], and the <size> vowel between [ɐ̄i], [ɐ̞̄i], and [āi]. The learner will therefore create different cue constraints for vowel quality, and these will be sensitive to the voicing of the following consonant (just as the cue constraints for vowel duration):

(50) *Cue constraints for vowel quality created by a learner of generation 2*

(/ɐi/[ai−vce])	/ɐi/[ai+vce]
(/ɐi/[ɐ̞i−vce])	/ɐi/[ɐ̞i+vce]
/ɐi/[ɐi−vce]	/ɐi/[ɐi+vce]
/ɐi/[ɐ̝i−vce]	(/ɐi/[ɐ̝i+vce])
/ɐi/[ʌi−vce]	(/ɐi/[ʌi+vce])

A computer simulation of the optimization of comprehension (as in Boersma 2006 and Boersma & Hamann 2007) shows that these constraints will be weighted partly by the peripherality bias, partly by their language-specific use: when compared to the rankings in §4.2, the weighting scale for the voiceless case is compressed, that for the voiced case expanded. With rounded figures, we get the weights in (51) and (52).

In <ice>, we now get an exaggerated raising:

(51) *Parallel production in generation 2*

	10	8	7	6	4	3	3	3	3	1	0	*H*
<ice>	<ice> |ɐis|	|ɐi| /ɐi/	/V/ [V̆−vce]	*[ɐ̞̆]	*[ɐ̆]	/ɐi/ [ɐ̞i −vce]	/ɐi/ [ɐi −vce]	/ɐi/ [ɐ̝i −vce]	/ɐi/ [ʌi −vce]	*[ɐ̝̆]	*[ʌ̆]	
☞ |ɐis| /ɐis/ [ʌ̆is]	+1	+1	+1						+1		−1	+28
|ɐis| /ɐis/ [ɐ̝̆is]	+1	+1	+1					+1		−1		+27
|ɐis| /ɐis/ [ɐ̆is]	+1	+1	+1		−1		+1					+24
|ɐis| /ɐis/ [ɐ̞̆is]	+1	+1	+1	−1		+1						+22

And in <size>, we get an exaggerated lowering:

(52) *Parallel production in generation 2*

	10	8	7	7	6	3	0	*H*
<size>	<size> |sɐjz|	|ɐj| /ɐj/	/V/ [V̄+vce]	/ɐi/ [ai +vce]	/ɐi/ [ɐ̞i +vce]	/ɐi/ [ɐi +vce]	/ɐi/ [ɐ̝i +vce]	
|sɐiz| /sɐiz/ [sɐ̝̄iz]	+1	+1	+1				+1	+25
|sɐiz| /sɐiz/ [sɐ̄iz]	+1	+1	+1			+1		+28
|sɐiz| /sɐiz/ [sɐ̞̄iz]	+1	+1	+1		+1			+31
☞ |sɐiz| /sɐiz/ [sāiz]	+1	+1	+1	+1				+32

Within two generations, then, the language has grown a large phonetic distinction: [ʌ̆i] before voiceless consonants, [ɐ̄i] before voiced consonants. (Comment: The process of enhancement will not continue catastrophically. In the present example it has led to the fairly large phonetic contrast [ʌ̆i]~[āi], but it might well have stopped short before had the parameters been slightly different. In general the cue constraints will force a halt before the auditory realization enters the realm of a neighbouring category, such as [ei] in the present case.)

4.4 Generation 3a creates a phonological contrast

For the speakers of generation 2, the contrast between [ʌ̆i] and [ɐ̄i] was ‘just phonetic’. But their children are confronted with a bigger contrast than the generation-2 speakers had been when young:

(53) *Language environment for a learner of generation 3*
<ice> [ʌ̆is]
<size> [sāiz]

The difference can be large enough for their children to reanalyse the distinction as categorical at the surface level, i.e. as the vowel categories /ʌi/ and /ai/. The learner will create cue constraints that take this contrast into account:

(54) *Cue constraints for vowel quality created by a learner of generation 3*

(/ʌi/[ai])	/ai/[ai]
(/ʌi/[ɐ̞i])	/ai/[ɐ̞i]
(/ʌi/[ɐi])	(/ai/[ɐi])
/ʌi/[ɐ̝i]	(/ai/[ɐ̝i])
/ʌi/[ʌi]	(/ai/[ʌi])

The cue constraints of this generation need no longer be conditioned by voicing.

In addition to the cue constraints, the listener can now also **create new structural constraints** for the cooccurrence of the features [lowered/raised] and [+vce/−vce] (Boersma, Escudero & Hayes 2003:1016; Hayes & Wilson 2006):

(55) *Structural constraints created by a learner of generation 3 (the same as in §2)*
/−vce/→/ʌi/
/+vce/→/ai/

Please notice that these constraints can be considered **natural**, hence **universal** constraints! Yet, the learner has created them from the language data with the help of her learning procedure (“innately guided empiricism”). What makes the constraints universal is that they ultimately (historically) derive from a universal phonetic bias.

Te next question is what the underlying form contains. It is possible that the surface level can contain major allophones that do not contrast lexically (Benders in prep.). If so, then the /ʌi/~/ai/ contrast can be restricted to the surface level, whereas the underlying form could still just contain |ɐi|, perhaps on the basis of alternations such as <life>~<life-s> [lʌ̆if]~[lāivz]. If this situation can be realistic, the ‘faithfulness’ constraints that the learner creates must have an arbitrary element:

(56) *‘Faithfulness’ constraints created by a learner of generation 3a*
|ɐi|/ʌi/
|ɐi|/ai/

The lexical constraints that generation 3a creates will be identical to those that their parents and grandparents created.

The choice between the allophones is now completely determined by structural constraints:

(57) *Parallel production in generation 3a*

	10	8	8	7	7	5	5	*H*
<ice>	<ice> |ɐis|	|ɐi| /ʌi/	|ɐi| /ai/	/−vce/→/ʌi/	/+vce/→/ai/	/ʌi/ [ʌi]	/ai/ [ai]	
☞ |ɐis| /ʌis/ [ʌ̆is]	+1	+1		+1		+1		+30
|ɐis| /ais/ [āis]	+1		+1				+1	+23

4.5 Generation 3b creates a lexical contrast

If alternations like <life>~<life-s> [lʌ̆if]~[lāivz] are unimportant, the lexicon receives no support for positing a single lexical phoneme. Therefore, the lexicon may start to contain |ʌi| and |ai|. The faithfulness constraints will be simplified:

(58) *Faithfulness constraints created by a learner of generation 3b*
|ʌi|/ʌi/
|ai|/ai/

The lexical constraints will now change as well:

(59) *Lexical constraints created by a learner of generation 3b*
<ice>|ʌis|
<size>|saiz|

Production proceeds as follows:

(60) *Parallel production in generation 3b*

<table>
<tr><td></td><td>10</td><td>8</td><td>8</td><td>7</td><td>7</td><td>5</td><td>5</td><td>H</td></tr>
<tr><td><ice></td><td><ice>
|ʌis|</td><td>|ʌi|
/ʌi/</td><td>|ai|
/ai/</td><td>/−vce/→/ʌi/</td><td>/+vce/→/ai/</td><td>/ʌi/
[ʌi]</td><td>/ai/
[ai]</td><td></td></tr>
<tr><td>☞ |ʌis| /ʌis/ [ʌ̆is]</td><td>+1</td><td>+1</td><td></td><td>+1</td><td></td><td>+1</td><td></td><td>+30</td></tr>
<tr><td>|ʌis| /ʌis/ [āis]</td><td>+1</td><td>+1</td><td></td><td>+1</td><td></td><td></td><td></td><td>+25</td></tr>
<tr><td>|ʌis| /ais/ [āis]</td><td>+1</td><td></td><td></td><td></td><td></td><td></td><td>+1</td><td>+15</td></tr>
</table>

This is just the story of how small phonetic learning biases can grow and then percolate up into the phonology and the lexicon: a story about phoneticization, phonologization, and lexicalization. It works in OT and in HG, and relies on the emergence of articulatory, cue, structural, faithfulness, and lexical constraints from the language input; without this emergence, the story would have been much more complicated.

A point to take home from §4 (in case you had not heard it yet from Boersma & Hamann 2007 or Boersma 2006b) is that the Bidirectional Phonology & Phonetics model solves the question (e.g. Hayes & Wilson 2006: 43) of how universally fixed phonetically-based scales of constraint rankings/weightings can exist when there are learning algorithms around that rank/weight the constraint in a way that is appropriate for the language. The answer is that, as in §4.3, computer simulations show that (cue or faithfulness) constraint rankings/weightings that have been learned from the ambient language while optimizing comprehension, may cause phonetically-looking biases when they are reused in production.

‘Some’ conclusions

Learning abstract URs:

(61) You may have noticed that the analysis of the interaction between Canadian raising and flapping did not involve learning underlying /t/ and /d/ for flap. This avoids the issue of how these neutralized segments are recovered by the learner.

We have not shown, or even suggested:

(62) ...that such recovery is in principle impossible, or that all cases can be reanalyzed along the lines we have proposed for this one case.

On natural constraints:

(63) You may have noticed that our learner is able to learn constraints that could produce patterns that probably exist in no human language.

An on-going line of research:

(64) Experimental studies of the learning natural and unnatural patterns, other studies of potential learning biases (Esper 1925, Schane *et al.* 1974, Finley yesterday), and computational modeling of these results (Wilson 2006)

Some possibilities based on our proposals:

(65) i. A more fully developed procedure for learning structural constraints might be used to model biases for simple featural generalizations

ii. Other biases might be in the phonetics itself (note that the model in section 4 renders moot the difficult question of how to incorporate phonetic biases into the phonology; cf. Hayes 1999, Wilson 2006, Flack 2007. This is also true in Flemming's 2001 HG work, but it is not clear how his single-level model can incorporate learning of unnatural phonotactic constraints)

Another ongoing line of research (section 4, Boersma 2006b, Boersma and Hamann 2007; see Kochetov 1999 Zuraw 2000, Wedel 2004 and others for related work):

(66) Explicit computational modeling of historical changes from neutral or unnatural patterns into natural ones

We have not shown, or even suggested:

(67) ...that doing phonological analysis, and making typological predictions, with an interacting set of natural constraints, is a bad idea. On the contrary, we think this is an excellent way of discovering important generalizations about language.

On phonologization and optimization:

(68) As far as we know, the proposal in section 4, is the only extant formally explicit analysis of the process of creating a phonological constraint or rule from a gradient pattern. Phonologization of gradient processes is often presented as an alternative to Optimality Theory (e.g. Blevins 2004). It's worth noting that this account of phonologization has at its heart an optimization-based choice procedure.

Some references

Albright, Adam, Giorgio Magri, Jennifer Michaels (2007). Paper to be presented at BUCLD.

Apoussidou, Diana (2007). *The learnability of metrical phonology.* PhD thesis, University of Amsterdam.

Bermúdez-Otero, Ricardo (2003). The acquisition of phonological opacity. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds), Variation within Optimality Theory: Proceedings of the Stockholm Workshop on 'Variation within Optimality Theory', 25–36. Stockholm: Department of Linguistics, Stockholm University. [Long version *ROA* **593**]

Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Boersma, Paul (1998). *Functional phonology*. PhD thesis, University of Amsterdam.

Boersma, Paul (2006a). Prototypicality judgments as inverted perception. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradedness in grammar*. Oxford: Oxford University Press. 167-184.

Boersma, Paul (2006b). The acquisition and evolution of faithfulness rankings. Talk at MFM 14, Manchester, May 27, 2006. [http://www.fon.hum.uva.nl/paul/]

Boersma, Paul (2006c). Parallel bidirectional phonology and phonetics. Talk at III Congresso Internacional de Fonética e Fonologia, Belo Horizonte, November 29, 2006. [http://www.fon.hum.uva.nl/paul/]

Boersma, Paul, Paola Escudero & Rachel Hayes (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*. 1013–1016.

Boersma, Paul, & Silke Hamann (2007). The evolution of auditory contrast. *ROA* **909**.

Boersma, Paul, & Joe Pater (2007). Convergence properties of a gradual learner for Harmonic Grammar. Ms.

Burzio, Luigi (2002). Surface-to-surface morphology: when your representations turn into constraints. In P. Boucher (ed.) *Many morphologies*. Cascadilla Press. 142–177. [Preliminary version *ROA* **341**]

Escudero, Paola (2005). *The attainment of optimal perception in second-language acquisition*. PhD thesis, Utrecht University.

Fischer, Markus (2005). *A Robbins-Monro type learning algorithm for an entropy maximizing version of stochastic Optimality Theory*. MA thesis, Humboldt University, Berlin. [*ROA* **767**]

Flack, Kathryn (2007). Inducing functionally grounded constraints. Ms. University of Massachusetts, Amherst. *ROA* **920**.

Goldsmith, John (1990). *Autosegmental and metrical phonology*. Malden: Blackwell.

Goldwater, Sharon, & Mark Johnson (2003). Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson, and Östen Dahl (eds.) *Proceedings of the Workshop on Variation within Optimality Theory*. Stockholm University. 111–120.

Idsardi, William (2006). Canadian raising, opacity and rephonemization. *Canadian Journal of Linguistics*.

Jäger, Gerhard (2003). Maximum entropy models and stochastic Optimality Theory. To appear in Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson & Annie Zaenen (eds.) *Architectures, rules, and preferences: a Festschrift for Joan Bresnan*. Stanford: CSLI. [available as *Rutgers Optimality Archive* **625**]

Legendre, Géraldine, Yoshiro Miyata, & Paul Smolensky. 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*, 884–891. Cambridge, MA: Lawrence Erlbaum.

Magri, Giorgio. In prep. MIT evals paper.

Moreton, Elliott, & Erik R. Thomas (2004). Origins of Canadian raising in voiceless-coda effects: a case study in phonologization. To appear in Jennifer Cole & José I. Hualde (eds.) *Papers in Laboratory Phonology 9*, Mouton-De Gruyter, Berlin.

Pater, Joe (1999). Austronesian nasal substitution and other NC effects. The prosody-morphology interface. CUP.

Pater, Joe (2007). Gradual learning and convergence. *Linguistic Inquiry*.

Pater, Joe, Rajesh Bhatt and Christopher Potts (2007). Linguistic optimization. Ms, University of Massachusetts, Amherst.

Smolensky, Paul, and Geraldine Legendre. 2006. *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.

Soderstrom, Melanie, Donald Mathis, & Paul Smolensky (2006). Abstract genomic encoding of Universal Grammar in Optimality Theory. In Paul Smolensky & Géraldine Legendre (eds.) *The harmonic mind.* MIT Press. 403–471.

Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* **30**: 945–982.