

# Learning Underlying Representations with Positive M-Phon Constraints

Diana Apoussidou, Karen Jesney and Joe Pater

*University of Massachusetts, Amherst*

{dapou1, kjesney, pater}@linguist.umass.edu

February 6<sup>th</sup>, 2008, Brown University

## 1. M-Phon constraints vs. abstract URs

A type of alternation in the phonological form of morphemes seen in many languages:

- (1) [mat] “bush”  
[mada] “bushes”

Assuming the plural morpheme is /-a/, “bush” has two forms:

- (2) [mat] [mad]

A hidden structure problem - which is the underlying (lexical) form?

- (3) /mat/ or /mad/?

Each hypothesis about the underlying form comes with a hypothesis about the rest of the language.

(4) /mat/      /mat+a/  
                 mada            *Intervocalic voicing*  
         [mat]      [mada]

This hypothesis is consistent with a language in which all intervocalic consonants are voiced, and final consonants are voiced and voiceless. Other words in the language could be:

(5) [bad]      [pada]

But not:

(6) \*[pata]

(7) /mad/     /mad+a/

mat

*Final devoicing*

[mat]     [mada]

This hypothesis is consistent with a language in which all final consonants are voiceless, and intervocalic consonants can be voiced and voiceless. For example:

(8) [bat]     [pata]

But not:

(9) \*[bad]

We build on Apoussidou's (2007, 2008) approach to this sort of problem, cast in the constraint-based Optimality Theory (OT: Prince and Smolensky 1993/2004).

We add three new ingredients:

- (10) i. Positive constraints induced by the learner (see also Boersma and Pater 2007, Pater et al. 2008)
  
- ii. Weighted constraints replace ranked ones (for learning advantages, see Johnson 2002, Goldwater and Johnson 2003, Boersma and Pater 2008)
  
- iii. Bias for Markedness > Faithfulness (proposed in OT by Smolensky 1996, see Jesney and Tessier 2007, 2008 in HG)

We first illustrate the proposal for a final devoicing language like Dutch (as in Apoussidou 2008; see Eisenstat 2008 for a very similar approach)



Markedness constraints (taken as given; could be induced)

(13) FINAL[-V]  
Word/syllable-final obstruents are voiceless

INTER[+V]  
Intervocalic obstruents are voiced

Faithfulness constraint:

(14) IDENT  
Underlying and surface values for voicing are identical

We adopt Noisy Exponential Harmonic Grammar (Boersma and Pater 2008). As a simple illustration of how this grammar model works, we can see how it chooses between [bad] and [bat] as a SR for UR /bad/.

Given the constraint weights...

$$(15) \quad \text{FINAL}[-V] = 0.935 \quad \text{IDENT} = 0.138$$

...noise is added... (sampled from a Gaussian distribution,  $\mu=0$ ,  $SD=0.2$ )


$$(16) \quad \text{FINAL}[-V] = 1.050 \quad \text{IDENT} = 0.160$$

...the exponential is taken... (to keep the weights positive)

$$(17) \quad \text{FINAL}[-V] = 2.858 \quad \text{IDENT} = 1.173$$

...and these weights are used to evaluate candidate mappings (constraint violations are negative integers, a candidate's score is the weighted sum of violations, and highest score wins)

(18)

	FINAL[-V]	IDENT	
/bad/	2.858	1.173	
[bad]	-1		-2.858
 [bat]		-1	-1.173

Learning uses Robust Interpretive Parsing (Tesar 1998)

- (19) The learner sees the overt forms only, and adds the optimal hidden structure based on the current state of the grammar. The resulting full structure is the “correct” form for error-driven learning.

The update procedure is HG-GLA (Boersma and Pater 2008) another name for {the Perceptron update rule, stochastic gradient ascent/descent, the Delta rule}.

The parameters (M>F bias imposed as in Jesney and Tessier 2008)

- (20) *Initial weights*

1 for Markedness constraints, 0 for others

*Learning rate*

0.01 for Markedness constraints, 0.001 for others

*Number of learning data*

100,000

A resulting grammar, shown with  $\exp(\text{weight})$  values:

(21)	FINAL[-V]	2.55
	/mad/	1.24
	IDENT	1.15
	/pat/	0.94
	/bad/	0.94
	/mat/	0.71
	INTER[+V]	0.44

Multiple runs yield similar results.

The tableaux now show candidate UR-SR pairs, and the proportion of evaluations in which each mapping is optimal (probability estimated by sampling)

(22)	“bush”	FINAL[-V] 2.55	/mad/ 1.24	IDENT 1.15	/mat/ 0.71	
0.95	/mat/ [mat]		-1			-1.24
0.05	/mad/ [mat]			-1	-1	-1.86
<0.0001	/mad/ [mad]	-1			-1	-3.26


(23)	“bushes”	/mad/ 1.24	IDENT 1.15	/mat/ 0.71	INTER[+V] 0.44	
<0.0001	/mat+a/ [mata]	-1			-1	-1.68
<0.00001	/mat+a/ [mada]	-1	-1			-2.39
1	/mad+a/ [mada]			-1		-0.71

The learner has arrived at a non-standard analysis – a different UR is (usually) chosen in each phonological context

- (24) /mat/ → [mat] “bush”  
 /mad+a/ → [mada] “bushes”

However, because of the  $M > F$  bias, the learner still does map unseen /bad/ to [bat]. Here is the same tableau as above, with mean weights and proportions.

(25)

		FINAL[-V]	IDENT	
	/bad/	2.55	1.15	
<0.01	[bad]	-1		-2.55
1	 [bat]		-1	-1.15

The learner passes the OT “richness of the base” test: the final state grammar filters out structures that are disallowed in the target language.

In other words, it generalizes appropriately (nonce word judgments, loanword pronunciations, and second language learner behavior can be used as evidence that humans generalize in this way)

A fairly substantial body of work has argued for versions of OT grammars that can choose URs based on the phonological context. This is termed allomorphy in the OT literature (see Wolf 2007 for an overview, and Apoussidou 2008 on learning allomorphy with M-Phon constraints)

Kager (2009) points out that allomorphy allows an alternative account of facts sometimes attributed to abstract URs. One of his examples is that of voicing alternations in Turkish. Our toy language becomes Turkish-like with the addition of another word.

(26)	[mat]	“bush”	[mada]	“bushes”
	[pat]	“tree”	[pata]	“trees”
	[bad]	“flower”	[bada]	“flowers”

This language has the [mat]/[mada] alternation, yet it permits final voiced consonants ([bad]) and intervocalic voiceless ones ([pata]).

There are essentially three types of consonant – the contrast is sometimes captured with an “archiphoneme”, that is, with underspecification of voicing (Inkelas, Orgun and Zoll 1997)

(27)	Fixed voiceless /t/	[-voice]
	Fixed voiced /d/	[+voice]
	Alternating /T/	[0voice]

With M-Phon constraints, the Turkish-type language can be acquired as a ranking/weighting of the same constraints as “Dutch”. The learner does not need to find the abstract UR.

Grammar learned for “Turkish”:

(28)	IDENT	1.69
	/mad/	0.99
	/pat/	0.97
	/bad/	0.97
	/mat/	0.95
	FINAL[-V]	0.83
	INTER[+V]	0.79

The high weight of IDENT allows contrast in all environments, including word-final position, as shown in the following table.

(29)

	“flower”	IDENT 1.69	/bad/ 0.97	FINAL[-V] 0.83	
0.01	/bad/ [bat]	-1			-1.69
0.99	/bad/ [bad]			-1	-0.83

But the markedness constraints can still choose between the URs for /mad/ ~ /mat/, thus yielding the alternation.

The following tableau shows how INTER[+V] chooses /mad/ when it is before a vowel – the choice of /mat/ in final position follows from a similar constraint interaction with FINAL[-V]

(30)	“bushes”	IDENT 1.69	/mad/ 0.99	/mat/ 0.95	INTER[+V] 0.79	
<0.005	/mat+a/ [mata]		-1		-1	-1.78
<0.00001	/mat+a/ [mada]	-1	-1			-2.68
1	/mad+a/ [mada]			-1		-0.95

The high weight of IDENT rules out /mat+a/ → [mada].

The constraints on URs have roughly equivalent weight, but INTER[+V] prefers the candidate with /mad/ in the UR, as in /mad+a/ → [mada].

Tesar (2006 CogSci, 2008 NECPHON) proposes a theory of UR learning that aims to find abstract URs like Turkish /T/. His toy example is more complicated than the Turkish one, but submits to a similar treatment.

	/re-/	/ri:-/	/'ro-/	/'ru:-/
/-se/	'rese	'ri:se	'rose	'ru:se
/-'si/	re'si	ri'si	'rosi	'ru:si
/-'so:/	re'so:	ri'so:	'roso	'ru:so

Tesar's "paka" language: stressed syllables preceded by single quote (e.g. 'ro), long vowels indicated by colon (e.g. ri:).

The apparent need for an abstract UR:

- (31) /ri:-/ must be underlyingly *long* to distinguish it from /re-/, and must be underlyingly *stressless* to distinguish it from /'ru:-/. In SRs it is *short unstressed* [ri] or *long stressed* ['ri:].

Our simulation used Tesar's Markedness and Faithfulness constraints, along with the M-Phon constraints constructed on the basis of the surface forms.

Using the same learning procedure as above, the learner always converges on a "correct" grammar for the observed forms (extremely low probability to other candidates), and the final state grammar also passes the richness of the base/generalization test.

The /ri-/ morpheme is distinguished in the structure of the constraint set (and by the constraint weights).

(32) *M-Phon constraints for "roots" in (31)*

*/'re/ /re/ /'ri:/ /ri/ /'ro/ /'ru:/*

There are in fact data that can be handled by M-Phon constraints that escape abstract URs.

French “schwa” is classic example of a case for which phonologists often posit an abstract segment. A word like *semaine* or *melon* can optionally be pronounced without the first vowel, while the phonetically identical vowel in *belon* cannot undergo deletion. French also has consonant clusters with no intervening vowel, as in *blonde* or *SMIC*.

- (33) *la s(e)maine*      ‘the week’  
*le m(e)lon*          ‘the melon’  
*la belon*            ‘the oyster (a particular kind)’  
*la blonde*          ‘the blonde’  
*le SMIC*            ‘the unemployment insurance’

Analysis in terms of abstract UR:

- (34) Alternating vowel: “V” – underspecified vowel  
Fixed vowel: /ə/ (or /œ/, /ø/) – fully specified vowel

This analysis is empirically inadequate. Not only is there a distinction between words that alternate and those that don't, but words that alternate differ in terms of how probable the pronunciation without the vowel is (see Pater 2008 on this assertion)

- (35) *le s(e)mestre*                      low probability of deletion  
*la s(e)maine*                              high probability of deletion

M-Phon constraints can be used to encode this sort of fine-grained lexical conditioning on the probability that a phonological process will apply.

Here we simply show that they can encode probability distributions over URs, and abstract from context (i.e. there are no markedness constraints).

Probabilities of each form in training data:

(36) *belon* 1  
      *s'maine* 0.7      *semaine* 0.3  
      *semestre* 0.6    *s'mestre* 0.4

Constraints (spelling instead of IPA):

(37) */belon/*  
      */s'maine/ /semaine/*  
      */s'mestre/ /semestre/*

Example resulting weights (training parameters as above)

(38)	/s'maine/	1.08
	/semestre/	1.04
	/belon/	1
	/s'mestre/	0.96
	/semaine/	0.93

Probability distributions produced by the grammar

(39)	belon	1	
	s'maine	0.70	semaine 0.30
	semestre	0.61	s'mestre 0.39

## Conclusions of Part 1

1. Positive M-Phon constraints can be used for representing *and* learning systems that might be thought to require abstract URs, and also for systems that abstract URs cannot handle at all.
2. If the number of observed M-Phon constraints is equal to the set of surface alternants, then there should be no concerns about the size of the constraint set/search space (cf. Tesar 2008)
3. If abstract URs are needed for other cases (as has been argued in the phonology literature), expanding the set of M-Phon constraints in the requisite ways seems tractable

## 2. Biases and learning with M-Phon constraints

Allomorphic analyses of systems like the Turkish one rely on faithful mappings between the UR and the surface.

- When the target form is non-alternating, only one UR is available. That UR maps faithfully.

(40) non-alternating form: “tree” “trees”  
UR selection /pat/ /pat+a/  
Surface realization: [pat] [pata]

(41) non-alternating form: “flower” “flowers”  
UR selection /bad/ /bad+a/  
Surface realization: [bad] [bada]

- IDENT is valued more highly than either FINAL[-V] or INTER[+V].
  - $w_{IDENT} > w_{FINAL[-V]}, INTER[+V]$
  - With only one underlying form possible / strongly preferred by the M-Phon constraints, that UR will be selected and mapped faithfully.
- When the target form alternates, more than one UR is available. The UR that can map faithfully to the unmarked output, is selected.

(42)	alternating form:	“bush”	“bushes”
	UR selection	/mat/	/mad+a/
	Surface realization:	[mat]	[mada]

But what about languages where voiced and voiceless consonants are not allowed in all contexts?

e.g., Dutch → final devoicing applies across the board

... this pattern should be independent of the UR selected

... we want our Dutch learner to learn that final consonants are predictably voiceless

The problem:

- Imagine you are a Dutch speaker exposed to the new word [bada], which you see only in its derived form.
- You recognize [-a] as an affix, so you segment out /bad/ and assign it the meaning “flower”.

- /bad/ is the only observed UR for “flower” and therefore the only one available.

## **What will happen if you are asked to produce the singular “flower” based on the UR /bad/?**

- Possibility 1 (the problem):
  - map /bad/ faithfully to [bad]
  - if  $w_{IDENT} > w_{FINAL[-V]}$ ... as in the full-contrast Turkish grammar
- Possibility 2 (the ideal):
  - map /bad/ unfaithfully to [bat]
  - requires  $w_{FINAL[-V]} > w_{IDENT}$

We cannot ensure that  $w_{\text{FINAL}[-V]} > w_{\text{IDENT}}$  if all constraints start with the same value.

Simulation: Target language contains two paradigms, both compatible with strict final devoicing:

{[pat], [pata]} *and* {[mat], [mada]}

All constraints begin with a value of 1.

Result: The target grammar is learned

(43)	"tree"	IDENT 1.80	FINAL[-V] 1.31	/pat/ 0.84	
1.00	/pat/ [pat]				0.00
0.00	/pat/ [pad]	-1	-1		-3.11

(44)	“trees”	IDENT 1.80	/pat/ 0.84	INTER[+V] 0.70	
>0.99	/pat+a/ [pata]			-1	-0.70
<0.01	/pat+a/ [pada]	-1			-1.80

(45)	“bush”	IDENT 1.80	FINAL[-V] 1.31	/mad/ 0.96	/mat/ 0.75	
>0.99	/mat/ [mat]		-1			-1.31
<0.01	/mad/ [mat]	-1			-1	-2.55

(46)	“bushes”	IDENT 1.80	/mad/ 0.96	/mat/ 0.75	INTER[+V] 0.70	
<0.01	/mat+a/ [mata]		-1		-1	-1.66
0.00	/mad+a/ [mata]	-1			-1	-2.50
0.00	/mat+a/ [mada]	-1	-1			-2.76
>0.99	/mad+a/ [mada]			-1		-0.75

But  $w_{IDENT} > w_{FINAL[-V]}, w_{INTER[+V]}$ .

So, the novel form “flower” will incorrectly map /bad/ → \*[bad] (rather than /bad/ → [bat]) on approximately 87% of trials.

The learned grammar is **unrestrictive**.

Beginning with  $M > F$  prevents this problem (see Smolensky 1996).

- Markedness constraints begin at 1, with a plasticity of 0.01
- Faithfulness and M-Phon constraints begin at 0, with a plasticity of 0.001

Result: the target grammar is learned restrictively

(47)	“tree”	FINAL[-V] 2.39	IDENT 1.17	/pat/ 0.88	
1.00	/pat/ [pat]				0.00
0.00	/pat/ [pad]	-1	-1		-3.56

(48)	“trees”	IDENT 1.17	/pat/ 0.88	INTER[+V] 0.52	
>0.99	/pat+a/ [pata]			-1	-0.52
<0.01	/pat+a/ [pada]	-1			-1.17

(49)	“bush”	FINAL[-V] 2.39	IDENT 1.17	/mad/ 1.00	/mat/ 0.77	
>0.99	/mat/ [mat]		-1			-1.17
<0.01	/mad/ [mat]	-1			-1	-3.16

(50)	“bushes”	IDENT 1.16	/mad/ 1.00	/mat/ 0.77	INTER[+V] 0.52	
<0.01	/mat+a/ [mata]		-1		-1	-1.52
0.00	/mad+a/ [mata]	-1			-1	-1.68
0.00	/mat+a/ [mada]	-1	-1			-2.16
>0.99	/mad+a/ [mada]			-1		-0.77

The preferred mapping here is still the faithful one – when possible.

- The UR /mat/ is preferred for /mat/→[mat]
- The UR /mad/ is preferred for /mad+a/→[mada]

In this grammar,  $w_{\text{FINAL}[-V]} > w_{\text{IDENT}}$

- Even if we've only seen [bada] – and so have no /bat/ constraint – the bare form will be realized as [bat] on over 99% of trials.

If M-Phon constraints are instead treated like Markedness constraints (i.e., beginning with a higher initial value) we get an even weaker preference for faithfulness:

(51)	“bush”	/mad/ 2.29	FINAL[-V] 1.72	/mat/ 0.97	IDENT 0.67	
0.10	/mat/ [mat]	-1				-2.29
0.90	/mad/ [mat]			-1	-1	-1.64

(52)	“bushes”	/mad/ 2.29	/mat/ 0.97	IDENT 0.67	INTER[+V] 0.26	
<0.01	/mat+a/ [mata]		-1		-1	-1.23
0.00	/mad+a/ [mata]	-1			-1	-2.55
0.00	/mat+a/ [mada]	-1	-1			-3.26
>0.99	/mad+a/ [mada]			-1		-0.67

Like the system with the  $M > F$ , M-Phon bias, this grammar also consistently maps the new input /bad/ to [bat].

... but the analysis more closely resembles the typical generative approach to this type of pattern.

If constraints are instead ranked – as in Optimality Theory (Prince & Smolensky 1993/2004) – the situation is a bit different.

As with HG, when all constraints begin at the same level (i.e., with no bias):

- the grammar prefers faithful mappings,
- ... the Markedness constraint FINAL[-V] ends up *slightly above* IDENT
- ... and the grammar is not fully restrictive. In the absence of a /bat/ constraint, ungrammatical [bad] will surface:

/bad/	→	*[bad]	0.13
		→ [bat]	0.87

Imposing the same  $M > F$  bias as with the HG learner resolves this issue – the grammar is fully restrictive.

/bad/ → \*[bad] 0.00  
→ [bat] 1.00

However, the OT learner with ranked constraints additionally requires that the M-Phon constraints be treated on par with the Faithfulness constraints.

- Both must begin with the same value. If not, it fails to converge on the correct grammar for the learning data.

M-Phon constraints, FINAL[-V] >> IDENT >> INTER[+V]  
100                      100                      50.51                      49.49

/pat/ → [pat] 1.00	/pat+a/ → [pata] 0.64
→ *[pad] 0.00	→ *[pada] 0.36

/mad/ → [mat] 0.50	/mad+a/ → [mada] 0.50
/mat/ → [mat] 0.50	/mat+a/ → [mada] 0.18
	→ *[mata] 0.32

The Tesar (2006) case discussed earlier has similar properties.

The target grammar admits only four forms:

### Attested forms

CVcv CV:cv

cvCV cvCV:

### Unattested forms

CVcv: CV:cv:

cv:CV cv:CV:

## Constraints:

MAINSTRESSLEFT – stress the leftmost syllable

MAINSTRESSRIGHT – stress the rightmost syllable

WEIGHTTOSTRESS – long vowels must be stressed

\*V: – don't have long vowels

IDENTSTRESS – output vowels should preserve input stress

IDENTLENGTH – output vowels should preserve input length

M-Phon constraints – /'re/ /re/ /'ri:/ /ri/ /'ro/ /'ru:/  
/se/ /'si/ /si/ /'so:/ /so/

With weighted constraints, the attested outputs are consistently selected for the training data  $>0.99$  over multiple trials.

With a Markedness  $>$  Faithfulness bias, the grammar is also restrictive.

**Observed inputs (=4)**

CVcv CV:cv

cvCV cvCV:

**Unobserved inputs (=12)**

CVcv: CV:cv:

cv:CV cv:CV:

CVCV: CV:CV

cv:cv CVCV:

CV:CV: CVCV

CV:CV: CVCV

- Only the four observed outputs are available.
- Inputs that are not consistent with the target language are mapped unfaithfully to outputs that conform to the language's phonotactics.
- Even on the more complex system, a restrictive grammar is learned when the appropriate bias is in place.

However - If no bias is enforced and all constraints start with the same value, however, the final state grammar will generalize inappropriately: it will accept structures not seen in the learning data.

/cv:CV/ → \*[cv:CV] ~ 0.27

/cv:CV:/ → \*[cv:CV:] ~ 0.27

/CV:CV:/ → \*[cv:CV:] ~ 0.23

Again, with ranked OT constraints, the correct outputs are consistently selected only when M-Phon constraints have the same initial value as Faithfulness constraints.

### **Summary:**

- HG biases: Markedness (M-Phon) > Faithfulness (M-Phon)
- OT biases: Markedness >> Faithfulness, M-Phon

### 3. Segmentation with M-Phon constraints

Consider a case where the learners have to find out the correct segmentation of words, as in Tibetan numerals:

(53) [ju] + [shi] = [jubshi]  
“ten” + “four” = “fourteen”

(54) [shi] + [ju] = [shibju]  
“four” + “ten” = “fourty”

Segmentation:

(55) /jub+shi/ or /ju+bshi/?  
/shib+ju/ or /shi+bju/?

Hypothesis 1: morpheme boundary is between word-internal consonant cluster

(56) /jub/, /shib/            /jub+shib/, /shib+jub/  
      [ju], [shi]                [jubshi], [shibju]

Grammatical consequence: word-final consonants delete on the surface.

But: inconsistent with other forms of the language:

(57) [jig] “one” ends in a consonant

Hypothesis 2: morpheme boundary is before word-internal consonant cluster

(58) /bju/, /bshi/      /bju+bshi/, /bshi+bju/  
[ju], [shi]      [jubshi], [shibju]

Grammatical consequence: word-initial consonant clusters are forbidden and resolves: /bju/ → #[ju].

How do the learners figure out that the segmentation is /ju+bshi/ and not /jub+shi/?

How does a learner find out that [ju] is underlyingly /bju/?

(59) Observed surface forms for Tibetan numerals:

[jig] “one”	[jugjig] “eleven”	
[shi] “four”	[jubshi] “fourteen”	[shibju] “fourty”
[na] “five”	[juna] “fifteen”	[nabju] “fifty”
[gu] “nine”	[jurgu] “nineteen”	[gubju] “ninety”
[ju] “ten”		

(60) M-Phon constraints:

{/jig/, /gjig/} “one”  
{/shi/, /bshi/, /shib/} “four”  
{/na/, /nab/} “five”  
{/gu/, /rgu/, /gub/} “nine”  
{/ju/, /bju/, /jub/, /jug/, /jur/} “ten”

Markedness constraint:

(61) \*#CC


No consonant clusters word-initially

Faithfulness constraint:

(62) MAX-C

Consonants in the UR are not deleted on the surface

A standard OT analysis:

/bju+gjig/	*#CC	MAX
[bjugjig]	*!	
 [jugjig]		*
[jujig]		**!

# Modeling segmentation of Tibetan numerals with Stochastic OT/GLA:

The parameters (no M>F bias)

(63) OTMulti in Praat

*Initial ranking values*

All equal (at 0)

*Learning rate*

1.0 (with a plasticity decrement of 0.1)

*Number of learning data*

400,000


*Noise*

2.0

A resulting grammar with ranking values:

(64)	“ten” /bju/	10.181
	“nine” /rgu/	8.868
	“four” /bshi/	8.309
	“five” /na/	5.498
	“one” /gjig/	3.686
	*#CC	1.975
	“ten” /ju/	-0.095
	:	
	:	
	“nine” /gu/	-6.179
	MAX-C	-18.749

Again, the tableau shows pairs of UR and SR pairs as candidates:

	“fourty”	/bju/	/bshi/	*#CC	/ju/	/shi/	MAX-C
0.00	/shib+ju/ [shibju]	*!	*			*	
>0.00	/shi+bju/ [shibju]		*!		*		
<1.00	 /bshi+bju/ [shibju]				*	*	*
0.00	/bshi+bju/ [bshibju]			*!	*	*	


## Resulting output distribution (of 10 learners):

/jig/ [jig]	>0.07	/bjubshu/ [jubshu]	<1.00
/gjig/ [jig]	<0.93		
		/bjubnu/ [junu]	<1.00
/shu/ [shu]	<0.01		
/bshu/ [shu]	>0.99	/bjubrgu/ [jurgu]	<1.00
/nu/ [nu]	<1.00	/shu+bjub/ [shubju]	<0.01
		/bshu+bjub/ [shubju]	>0.99
/gu/ [gu]	<0.01		
/rgu/ [gu]	>0.99	/nu+bjub/ [nubju]	<1.00
/bjub/	<1.00	/gu+bjub/ [gubju]	<0.01
/bjub+gjig/ [jugjig]	<1.00	/rgu+bjub/ [gubju]	>0.99

Segmentation of morphemes and grammar are successfully learned: on the surface, no word-initial clusters occur.

The interaction between the M-Phon constraints causes a minimal use of allomorphy.

The learners pass the OT “richness of the base” test: the final state grammar filters out structures that are disallowed in the target language, i.e. \*#CC-structures.

/gju+bji/	*#CC	MAX
[gjubji]	*!	
 [jubji]		*

Modeling segmentation of Tibetan numerals with exponential HG-GLA renders slightly different results.

The parameters (M>F bias imposed as in Jesney and Tessier 2008)

(53) *Initial weights*

1 for Markedness constraint, 0 for others

*Learning rate*

0.01 for Markedness constraint, 0.001 for others

*Number of learning data*

100,000

A resulting grammar, shown with exp(weight) values:

(65)	“ten” /bju/	1.153
	*#CC	0.953
	“four” /bshi/	0.673
	“nine” /rgu/	0.623
	“five” /na/	0.403
	“one” /gjig/	0.283
	“ten” /ju/	0.063
	“nine” /gu/	0.003
	“four” /shi/	-0.037
	MAX-C	-0.103
	{Rest of M-Phon}	<-0.1

A HG-learner using allomorphy:

(66)	“fourty”	/bju/	*#CC	/bshi/	/ju/	/shi/	MAX-C	
		3.17	2.59	1.96	1.06	0.96	0.9	
0.00	/shib+ju/ [shibju]	-1		-1		-1		-5.13
<0.47	/shi+bju/ [shibju]			-1	-1			-3.02
>0.53	/bshi+bju/ [shibju]				-1	-1	-1	-2.11
0.00	/bshi+bju/ [bshibju]		-1		-1	-1		-4.61

## Resulting output distribution:

/jig/ [jig]	>0.72	/ju+bshi/ [jubshi]	<0.03
/gjig/ [jig]	<0.28	/bjubshi/ [jubshi]	>0.97
/shi/ [shi]	<0.47	/ju+na+ [juna]	<0.03
/bshi/ [shi]	>0.53	/bjuna+ [juna]	>0.97
/na/ [na]	<1.00	/ju+rgu/ [jurgu]	<0.03
/gu/ [gu]	<0.50	/bjurgu+ [jurgu]	>0.97
/rgu/ [gu]	>0.50	/shi+uju/ [shibu]	<0.47
/ju/ [ju]	<0.03	/bshibu+ [shibu]	>0.53
/uju/ [ju]	<0.97	/na+uju/ [nabu]	<1.00
/ju+gjig/ [jugjig]	>0.03	/gu+uju/ [gubu]	<0.50
/uju+gjig/ [jugjig]	<0.98	/rgu+uju/ [gubu]	>0.50

Still, crucial weightings of constraints are maintained: all HG-learners have  $M > F$ , and rule out ungrammatical inputs (i.e. generalize correctly).

### **Summary of this section:**

- HG biases: Markedness  $>$  Faithfulness, M-Phon
- OT (no) biases: Markedness, Faithfulness, M-Phon

Issues:

- The candidate set
- The size of the M-Phon constraint set
- The learnability of M-Phon constraints

A simple hidden structure problem with a local minimum

Form A	C1	C2	C3
Overt-Correct (Parse-1)	-1		
Overt-Correct (Parse-2)		-1	
Incorrect			-1

Form B	C1	C2	C3
Correct			-1
Incorrect		-1	

Local minimum: C1 >> C2, C3

A simple hidden structure problem without (?) a local minimum

Form A	C1	C2	C3
Overt-Correct (Parse-1)	-1		
Overt-Correct (Parse-2)		-1	
Incorrect			-1

Form B	C1	C2	C3
Correct		-1	
Incorrect	-1		

No local minimum?

## The Dutch hidden structure problem

Form A	/mat/	/mad/	Inter[+V]	Ident
/mada/ → [mada]	-1			
/mata/ → [mada]		-1		-1
/mata/ → *[mata]		-1	-1	

Form B	/mat/	/mad/	Inter[+V]	Ident
/pata/ → [pata]			-1	
/pata/ → *[pada]				-1

OT requires: /mad/ >> /mat/; Ident >> Inter[+V]

HG allows: w/ident > w/inter[+V];

w/mad/ + w/inter[+V] > w/mat/

## The Dutch hidden structure problem

Form A	C1	C2	C3	C4
Overt-Correct 1	-1			
Overt-Correct 2		-1		-1
Incorrect		-1	-1	

Form B	C1	C2	C3	C4
Overt-Correct			-1	
Incorrect				-1

OT requires:  $C4 \gg C3$ ;  $C2 \gg C1$

HG allows:  $wC4 > wC3$ ;  
 $wC2 + wC3 > wC1$

## References

Apoussidou, Diana. 2007. The learnability of metrical phonology. Doctoral Dissertation, University of Amsterdam. Available at <http://roa.rutgers.edu/>.

Apoussidou, Diana. 2008. Final devoicing, allomorphy, and Freedom of the Lexicon: An online learning approach to underlying representations. Talk presented at Yale University, October 30<sup>th</sup>.

Boersma, Paul. 2001. Phonology-semantics interaction in OT, and its acquisition. In Robert Kirchner, Wolf Wikeley & Joe Pater (eds.): *Papers in Experimental and Theoretical Linguistics*, Vol. 6: 24-35. Edmonton: University of Alberta.

Boersma, Paul, and Joe Pater. 2007. Constructing constraints from language data: The case of Canadian English diphthongs. Paper presented to the North Eastern

Linguistic Society (handout available at <http://people.umass.edu/pater/boersma-pater-nels.pdf>).

Boersma, Paul, and Joe Pater. 2008. Convergence properties of a gradual learning algorithm for Harmonic Grammar. Ms, University of Amsterdam and UMass Amherst. Available at <http://roa.rutgers.edu/>.

Jesney, Karen and Anne-Michelle Tessier. 2008. Gradual learning and faithfulness: consequences of ranked vs. weighted constraints. To appear in Muhammad Abdurrahman, Anisa Schardl & Martin Walkow (eds.), *Proceedings of the 38th Meeting of the North East Linguistics Society (NELS 38)*. Available at <http://roa.rutgers.edu/>.

Kager, René. 2009. Lexical irregularity and the typology of contrast. In Kristin Hanson and Sharon Inkelas (eds.), *The*

Nature of the Word: Essays in Honor of Paul Kiparsky.  
Cambridge, MA: MIT Press. [http://www.let.uu.nl/  
~Rene.Kager/personal/Papers/contrast.pdf](http://www.let.uu.nl/~Rene.Kager/personal/Papers/contrast.pdf)

Pater, Joe. 2008. Lexically conditioned variation in Harmonic Grammar. Paper presented at OCP-5, Toulouse.  
<http://people.umass.edu/pater/pater-french-schwa.pdf>

Pater, Joe, Elliott Moreton and Michael Becker. 2008. Simplicity biases in structured statistical learning. Poster presented at BUCLD. <http://people.umass.edu/pater/pater-buclid.pdf>

Tesar, Bruce. 2006. Faithful contrastive features in learning. *Cognitive Science* 30:5, 863-903.  
<http://rucss.rutgers.edu/~tesar/papers/cogsci2006.pdf>

Tesar, Bruce. 1998. An iterative strategy for language learning. *Lingua* 104:131-145.

<http://rucss.rutgers.edu/~tesar/papers/lingua98.pdf>

Wolf, Matthew. 2008. Optimal Interleaving: Serial Phonology-Morphology Interacton in a Constraint-Based Model. Ph.D. dissertation, UMass Amherst. Available at <http://roa.rutgers.edu/>.