Empirical motivation for transparent interactions in the English lexicon

James Myers
National Chung Cheng University
Lngmyers@ccu.edu.tw

Overview

- Quantitative arguments from dictionary data in phonology
- Theoretical implications of this way of looking at phonological data
- Applications to some transparent interactions in English

Arguing from dictionary data

(1) Phonological argumentation is ultimately quantitative:

- Are [d] and [t] allophones in English?

  madam ten Teddy Toledo today ... Biased sample!

- Do /t/ and /d/ become flaps in English, or the other way around?

  á[ɾ]om ~ a[tʰ]ómic {t, d} → ɾ: no exceptions
  é[ɾ]it ~ e[d]ítion ɾ → d: *atom* is an exception; ɾ → t: *edit* is an exception

- Does English have a /k/ copying rule (Chomsky and Halle 1968, Borowsky 1990)?

  a<u>cc</u>ess, a<u>cc</u>ident, a<u>cc</u>ept, a<u>cc</u>elerate, a<u>cc</u>ede, a<u>cc</u>ent [ks], all from Latin *ad* + *c*...
  cf. su<u>cc</u>eed [ks] vs. su<u>sc</u>eptible [s] & su<u>cc</u>inct [s], all from Latin *sub* + *c*...

- Does English have constraints against /bn/ and /sf/?

  If no *bn constraint, /b/ & /n/ appear freely ... making total absence of /bn/ unlikely.
  If no *sf constraint, /s/ & /f/ appear freely ... making the rarity of /sf/ unlikely:
  sphere, *sphinx*... and that's it.

(2) Basic quantitative logic depends on relative type frequencies:

- For pattern P, count O = number(obey(P)) vs. V = number(violate(P)).
- O should be so much higher than V that the difference is unlikely to be due to chance.
- The ratio between V to O represents the strength of the pattern.

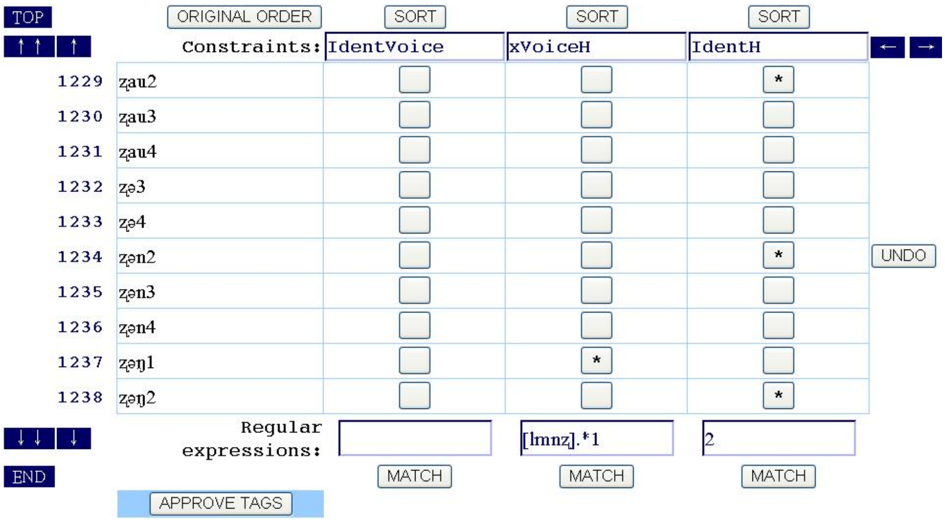
(3) Statistics can tell us whether a pattern's strength is so much better than the null hypothesis (O = V) that the pattern is unlikely to be due to chance. E.g. for /ks/ pattern:

- 7 words obey it (access, accident, accept, accelerate, accede, accent, succeed)
- 2 words violate it (susceptible, succinct)
- Chance would be like flipping a coin 9 times and getting 7 heads (or 7 tails).
- There are 2^9 = 512 possible ways to flip 9 coins; 92 of them show 7 heads or tails.
- So chance probability p = 92/512 = 0.1797.
- That goes over the p > .05 convention, so not statistically significant.

(4) In an output-constraint approach like Optimality Theory (OT), we can automatically mark constraint violations by searching for strings (or string patterns) that are banned by the constraint (easiest to do with markedness constraints).

*sf Mark all words containing the string ...sf...

(5) MiniCorp (Myers 2008, forthcoming) does this automatically.
<www.ccunix.ccu.edu.tw/~lngproc/MiniCorp.htm>



(6) The strength of an OT constraint is related to the notion of constraint weight, as in Harmonic Grammar (e.g. Legendre, Sorace, & Smolensky 2006). The constraint weights in a Harmonic Grammar can be set automatically from dictionary data by various learners (e.g. Hayes and Wilson 2008, Coetzee and Pater 2008).

(7) MiniCorp not only sets constraint weights, but also tests them for statistical significance (which other models don't do), by using a generalization of the coin-flipping logic.

- First count type frequencies of classes defined by all possible constraint evaluations:

Counts	**magical /ks/ constraint**	treated	**Output variable**	**Input variable**
7		as	7	0
2	*	...	2	1

- Then run a count-based (Poisson) weight-fitting (regression) model:

ks constraint weight = -1.25 (negative, so obeyed more than violated)
[Note that -1.25 = ln(2/7), where ln = natural logarithm, based on e = 2.718...]
ks significance: p = .118 (p > .05, so not strong enough to be better than chance)
[Model also gives a baseline weight of 1.95 = ln(7)]

- Counts can be predicted by multiplying weights by violations for each constraint (or baseline) and summing them (then raise e to the power of the total)

baseline + weight × 0 = 1.95 + (-1.25) × 0 = 1.95... and $e^{1.95} \approx 7$
baseline + weight × 1 = 1.95 + (-1.25) × 1 = 0.70... and $e^{0.70} \approx 2$

(8) OT is a special case of Harmonic Grammar, where constraint weights are restricted so that the weight of each constraint is larger than the sum of the weights (times maximum violations) for all lower-ranked constraints (Prince and Smolensky 2004, Prince 2007). This way, the optimal candidate will still have the lowest violation score:

input	C1	C2	C3
output1	*		
☞ output2		*	**

workable weights:	6	3	1
	C1	C2	C3
6 = 6×1 + 3×0 + 1×0	1	0	0
5 = 6×0 + 3×1 + 1×2	0	1	2

(9) MiniCorp can test whether this ranking condition is met (which other models don't do), by comparing two regression models. For example, for the simplest two-constraint case:

Counts = baseline + $weight_1$ × Constraint_1 + $weight_2$ × Constraint_2

Ranking model: Assume $weight_1 \neq weight_2$
No ranking model: Assume $weight_1 = weight_2$

(10) For example, consider the following the fake data:

Fake data

Counts	**C1**	**C2**
100		
50		*
40	*	
0	*	*

Results of MiniCorp-style analysis

Constraints	**Weights**	**p**
C1	-1.32	< .0001
C2	-1.03	< .0001
Ranking		**p**
C1 >> C2		.23

- Each constraint is statistically significant, and their weights differ in the desired way:

  C1 weight = ln((40+0)/(100+50)) = -1.32, $p < .0001$
  C2 weight = ln((50+0)/(100+40)) = -1.03, $p < .0001$
  |weight(C1)| > |weight(C2)|, which follows from 50 > 40

- But the constraint weight difference isn't big enough to be significant ($p = .23 > .05$). That's hinted at by the counts in the middle of the counts table: 50/40 = 1.25 ≈ 1.

(11) Myers (2008, forthcoming) uses MiniCorp to test various constraints and rankings in Mandarin phonotactics.

Theoretical implications

(12) Lexical exceptions are ungrammatical.
They are "acceptable" to speakers, because they know them despite the grammar, in some extragrammatical way (e.g. rote memory). Thus attempts to incorporate exceptions into the grammar (e.g. Pater forthcoming) miss the point.

(13) Patterns can differ in strength (i.e. V/O ratios).
This isn't the same as productivity: the probability of applying a pattern in novel forms (Myers 1993, forthcoming), which can only be tested in fluent corpora or experiments.

(14) Pattern strength is relevant to learnability, but the relationship is complex.

- Weak constraints may not be learned (even if babies don't care about *p* values).
- Yet babies presumably also have innate learning biases (analytic bias; Moreton 2008), imposing patterns that aren't in the input, or even ignoring patterns that are there.
- At best, pattern strength reflects how well grammar was implemented diachronically; the main interest of the type frequency logic is for the linguist, not baby-as-linguist.

(15) Even for the linguist, type frequency logic raises further puzzles. For example, which of the following two data sets gives stronger evidence for the constraint C?

Data set A

Counts	C	
5		Weight = -1.61
1	*	$p = .14$

Data set B

Counts	C	
10		Weight = -2.30
1	*	$p = .03$

- Why should increasing the number of items where C is *not* violated (i.e. the string defining C is *absent*) improve the strength of C?
- Reiss (2008) critiques constraint-based theories (not just OT) for this reason, citing nonsense constraints like NoBanana (no sentence may contain an actual banana).
- Yet similar problems beset all scientific logic, as in the raven paradox (Hempel 1945): Confidence in the statement "All ravens are black" is increased by finding a black raven. Since this statement is logically equivalent to "All non-black things are non-ravens", then we can support "All ravens are black" with a green apple.
- Hempel's solution was to say that green apples do indeed support the black raven statement, but it's only obvious if the universe is small. That's exactly the case with the above data sets. With data sets of more realistic sizes, there's no real difference:

Data set A'

Counts	C	
5000		Weight = -8.52
1	*	$p < .0001$

Data set B'

Counts	C	
10000		Weight = -9.21
1	*	$p < .0001$

(16) Another problem: If ranking is defined in terms of weights, and weights also reflect pattern strength, which reflects type frequencies, then we can get evidence for ranking directly from type frequencies, even if the constraints never interact within any word:

/UXU/	*X	*Y
[X]	*	
[Y]		*
☞ [Z]		

/UYU/	*X	*Y
[X]	*	
[Y]		*
☞ [Z]		

Counts	*X	*Y	Examples
100			ABA, ACA, ADA, ...
60		*	AYA, ...
40	*		AXA, ...
0	*	*	(none)

Constraints	Weights	p
*X	-1.39	< .0001
*Y	-0.85	< .0001
Ranking		**p**
*X >> *Y		.021

- But maybe this is an empirical question...? (See below.)

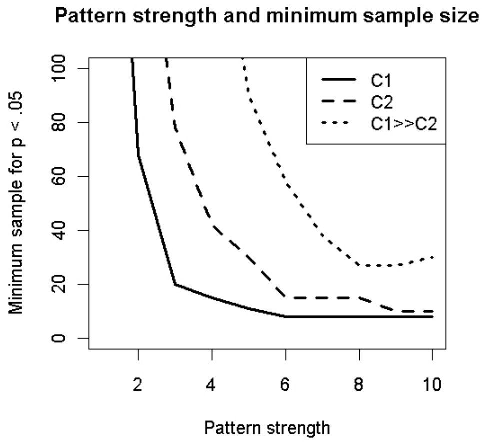
(17) Factors potentially influencing the researcher's ability to detect constraints and rankings:

- Pattern strength: easier detection with larger type frequency differences...?
- Ranking strength: easier detection with larger constraint weight differences...?
- Number of constraints: easier detection with fewer constraints...?
- Gradience: easier detection if constraints give at most one star...?

(18) The effect of pattern strength (simulated via baseline in a model with equal counts)

Pattern strength = 2

Counts	C1	C2
37		
24		*
20	*	
18	*	*



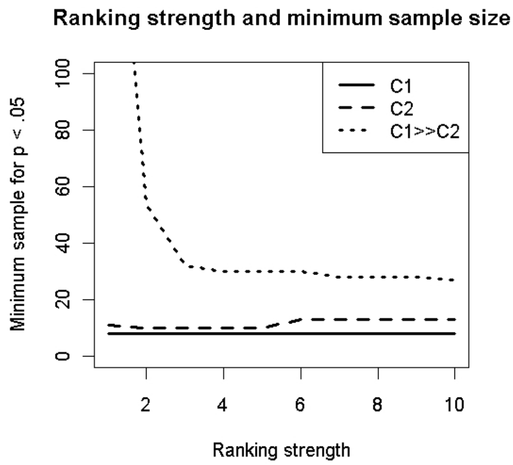
Pattern strength = 10

Counts	C1	C2
86		
23		*
0	*	
0	*	*

(19) The effect of ranking strength (simulated via minimum constraint weight difference)

ranking strength = 1

Counts	C1	C2
80		
17		*
10	*	
0	*	*



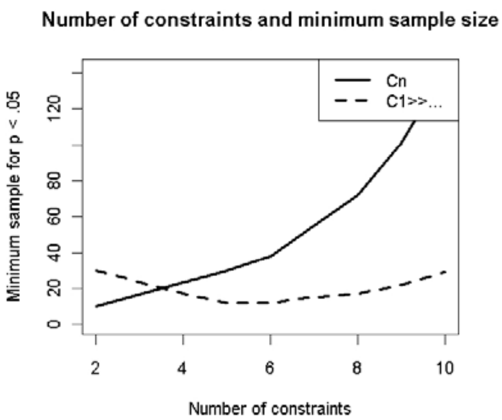
ranking strength = 10

Counts	C1	C2
89		
26		*
0	*	
0	*	*

(20) The effect of the number of constraints

Number of constraints = 2

Counts	C1	C2
86		
23		*
0	*	
0	*	*



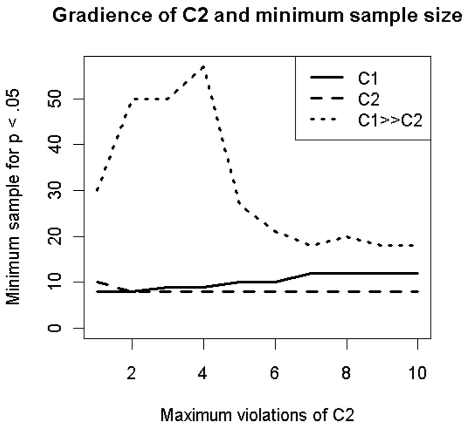
Number of constraints = 10

Counts	C1	...	C8	C9	C10
100		...			
37		...			*
14		...		*	
5		...		*	*
2		...	*		
1		...	*		*
0		...	*	*	
...	...	...	...	...	...
0	*	...	*	*	*

(21) The effect of constraint gradience (simulated via max violations of lower constraint)

max violations of C2 = 1

Counts	C1	C2
86		
23		*
0	*	
0	*	*



max violations of C2 = 10

Counts	C1	C2
97		
34		*
11		**
2		***
0		****
...	...	...
0		**********
0	*	
0	*	*
0	*	**
...	...	...
0	*	**********

Some transparent interactions in English

(22) Transparent interactions (Kiparsky 1976, McCarthy 2007):

- An interaction between patterns is transparent if both patterns are surface-true (no underapplication) and surface-apparent (no overapplication)
- Ordinary OT can do it (no serial derivation, only input-output correspondence)

(23) Some transparent rule interactions in English lexical phonology (Chomsky and Halle 1968, Rubach 1984, 1996, Halle and Mohanan 1985)

CiV lengthening	Caucasus [ə] ~ Caucasian [ej]
s-voicing	Caucasus [s] ~ Caucasian [ʒ]
i-shortening	Palestine [aj] ~ Palestinian [ɪ]
Prenasal g-deletion	signal [gn] ~ sign [n]
Compensatory lengthening	signal [ɪ] ~ sign [aj]
{Vowel Shift	/i:/ → [aj], /e:/ → [ij], /æ:/ → [ej]}
{Palatalization	confuse [z] ~ confusion [ʒ]}

(24) s-Voicing s → [+voiced] / V: _ V

- recluse [s] ~ reclusion [u:ʒə] {plus transparent interaction with palatalization}
- sign [s] ~ design [əzai] {plus opaque interaction with vowel reduction}
- music [u:zɪ] {ignoring morphological context, as SPE does}

(25) Rethinking this pattern in output-oriented terms:

Markedness: *V:SV where S represents a voiceless sibilant fricative (i.e. [s] or [ʃ])
Faithfulness: Ident(Voice)
*V:SV >> Ident(Voice)

reclu/s/+ion	*V:SV	Id(Voice)
[ʃ]	*	
☞ [ʒ]		*

(26) Note that s-voicing interacts opaquely with spirantization: {t, d} → {s, z} _ j

- relate ~ relation [ejʃə] (*[ejʒə]) (though cf. equate ~ equation [ejʒə])

(27) ... and with velar softening: k → s / _ {i, e}

- Greek ~ Grecian [ijʃə] (*[ijʒə])

(28) CiV lengthening V → V: / _ C ĭ V {i is unstressed}
i-shortening i: → i / _ C i V

- Iran [æ] ~ Iranian [ejnə] {plus transparent Vowel Shift}
- Palestine [aj] ~ Palestinian [ɪniə]

*VCvV {V = short, v = unstressed} (Alcántara 1998 analyzes this as mora relinking)
*i: >> *e: >> *æ: Length is linked to sonority (e.g. Alcántara 1998)
*i: >> *VCvV >> *e: >> *æ:

Ir/æ/n + ian	*i:	*VCvV	*æ:
[æ]		*	
☞ [ej]			*

Palest/aj/n + ian	*i:	*VCvV	*æ:
☞ [ɪ]		*	
[aj]	*		

(29) CiV lengthening feeds s-Voicing (i.e. helps creates the environment for it):

- Caucasus [əs] ~ Caucasian [ejʒ]: lengthen vowel, then voice /s/

But in OT there's no need to rank the relevant markedness constraints:

Cauca/ə/s + ian	*V:SV	*VCvV
[æʃ]		*
[ejʃ]	*	
[æʒ]		*
☞ [ejʒ]		

(30) Prenasal g-deletion

μ μ
| → / _ [nasal] $]_\sigma$
g

Compensatory lengthening

μ μ μ μ
| → | /
V V

- signal [ɪgn] ~ sign [ajn] paradigmatic [ɪgm] ~ paradigm [ajm] (feeding)

- In OT the processes need not interact: {Max(μ), $\boxed{\text{*g}^{\mu}\text{Nas]}_\sigma}$} >> Max(C); $\boxed{\text{Link-}\mu}$

- But if moras are predictable, they're not underlying, so what does Max(μ) do?
- This kind of interaction may actually be opaque (Shaw 2007)...?

(31) For convenience, follow Halle and Mohanan (1985:96): combine them into one pattern:

*V(g)n For input /gn/, ban short vowels before syllable-final nasal (ugh)

s/ig/n	*V(g)n
[ɪg]	*
[ajg]	*
[ɪ]	*
☞ [aj]	

s/ig/n	*V(g)n	*i:
[ɪg]	*	
[ajg]	*	*
[ɪ]	*	
☞ [aj]		*

Thus prenasal g-deletion forces violation of *i:

(32) So we end up with the following partial rankings:

*V:SV *{s-voicing}* >> Ident(Voice)
*V(g)n *{prenasal g-deletionl}* >> *i: *{i-shortening}* >> *VCvV *{CiV lengthening}* >> *e: ...

(33) Testing these rankings by counting word types in the CMU Pronouncing Dictionary (at Carnegie Mellon University; <www.speech.cs.cmu.edu/cgi-bin/cmudict >):

- North American English
- 127,008 words (version 0.6, which is what I used)
- Transcription uses the ASCII-based Arpabet:

```
ENGLISH  IH1 NG G L IH2 SH
PHONOLOGICAL  F OW2 N AH0 L AA1 JH IH0 K AH0 L
PATTERN  P AE1 T ER0 N
```

- Too big and messy for current version of MiniCorp, so I did the analyses with special-purpose scripts written in R (R Core Development Team 2008)

(34) Restricted analyses to 2016 words ending in the orthographic letter strings EOUS, IO, IAN, IAL, IA, IOUS, IUM, and GN, GM (relevant for prenasal g-deletion)

(35) Violations of markedness constraints defined by specified substrings:

*i:, *e: [aj] or [ij]
*VCvV short vowel - consonant - unstressed vowel - any vowel
*V:SV long vowel - {s, ʃ} - any vowel
*V(g)n short vowel - nasal - end of word

(36) To control for opacity, I defined violations of last two constraints within subsets:

*V:SV within words written with "S" or "C" (avoids spirantization cases like *relation*)
*V(g)n within words written with "GN" or "GM" (avoids no-g cases like *wine*, *win*)

(37) Violations of faith constraints are harder to mark automatically, so I only tried one:

Ident(Voice) [z] or [ʒ] within words written with "S" or "C"

(38) Initial test of *V:SV >> Ident(Voice)

Counts	*V:SV	Ident(Voice)	Examples
1866			
64		*	Caucasian [ʒ], gymnasium [z]; Keynesian [z]
86	*		Andalusian [ʃ]; racial [ʃ], Confucian [ʃ], spacious [ʃ]
0	*	*	

Constraints	Weights	p
*V:SV	-3.11	< .0001
Ident(Voice)	-3.42	< .0001
Ranking		**p**
*V:SV >> Ident(Voice)		NA

Significant despite the many exceptions, but...

...weight difference goes the wrong way!

(39) Most of the exceptions to *V:SV (63/86) are written with "C". If we eliminate these:

Counts	*V:SV	Ident(Voice)
1929		
64		*
23	*	
0	*	*

Constraints	Weights	p
*V:SV	-4.46	< .0001
Ident(Voice)	-3.42	< .0001
Ranking		**p**
*V:SV >> Ident(Voice)		< .0001

(40) But this isn't a case of opacity, since very few of these "C" words would be considered to have an underlying /k/ in modern English (cf. *Grecian*).

Halle & Mohanan (1985:99) consider them exceptions, just like any other. If this is right, the ranking of *V:SV relative to Ident(Voice) cannot be confirmed.

(41) Testing *i: >> *VCvV

Counts	*i:	*VCvV	Examples
1390			
623		*	Darwinian [ɪ]; librarian [ɛ], curious [ʊ]; gaseous [æ], millennium [ɛ], Yugoslavian [a], Orwellian [ɛ]
3	*		chlamydia [aj], Irian [aj], xylia [aj]
0	*	*	

Constraints	Weights	p
*i:	-6.51	< .0001
*VCvV	-0.80	< .0001
Ranking		**p**
*i: >> *VCvV		< .0001

(42) Testing *V(g)n >> *i:

Counts	*V(g)n	*i:	Examples
1911			
101		*	(anything with [aj] in it)
3	*		ensign [ɪn], foreign [ɪn], sovereign [ɪn]
1	*	*	diaphragm [æm]

(orthography ≠ real input...)

Constraints	Weights	p
*V(g)n	-6.22	< .0001
*i:	-2.93	< .0001
Ranking		**p**
*V(g)n >> *i:		< .0001

(43) Adding *V:SV... We don't expect *V:SV and *VCvV to be ranked (see (29) above)...

Counts	*i:	*V:SV	*VCvV
1304			
623			*
86		*	
0		*	*
3	*		
0	*		*
0	*	*	
0	*	*	*

Constraints	Weights	p
*i:	-6.51	< .001
*V:SV	-3.11	< .001
*VCvV	-0.80	< .001
Ranking		**p**
*i: >> {*V:SV, *VCvV}		< .001
*V:SV >> *VCvV		< .001

... But they are anyway. Is this a bad result for the type frequency approach...?

(44) Yet transitivity logic in ranking is paralleled by transitivity logic in type frequency:

*V(g)n >> *i: e.g. sign [aj]
*i: >> *VCvV e.g. Palestinian [ɪ]
Therefore: *V(g)n >> *i: >> *VCvV {no direct evidence for this full ranking}

Counts	*V(g)n	*i:	*VCvV
1301			
605			*
79		*	
18		*	*
8	*		
0	*		*
5	*	*	
0	*	*	*

Constraints	Weights	p
*V(g)n	-5.04	< .0001
*i:	-2.93	< .0001
*VCvV	-0.80	< .0001
Ranking		**p**
*V(g)n >> {*i:, *VCvV}		< .0001
*i: >> *VCvV		< .0001

(45) What about *i: >> *e:, inferred indirectly from *i: >> *VCvV >> *e: ...?

Counts	*i:	*e:
401		
1513		*
53	*	
49	*	*

② ... but other vowels are rarer than these two combined.

① Type frequencies differ in the expected way...

Support for *i: (53+49) < (401+1513): weight = -2.93 negative (good)
Support for *e: (1513+49) > (401+53): weight = 1.24 positive (bad!)

(46) Assuming that full corpus supports *e: (most vowels aren't [ij]), plug in a fake number:

Counts	*i:	*e:
2513[a]		
1513		*
53	*	
49	*	*

[a] (= max(Counts) + 1000; adjusted from 401)

Constraints	Weights	p
*i:	-3.68	< .0001
*e:	-0.50	< .0001
Ranking		**p**
*i: >> *e:		< .0001

(47) Does this ranking difference accommodate all the other constraints...?

*V(g)n >> *i: >> *VCvV >> *e: ...

Counts	*V(g)n	*i:	*VCvV	*e:
1913[a]				
913				*
5			*	
600			*	*
52		*		
31		*		*
0		*	*	
18		*	*	*
3	*			
0	*			*
0	*		*	
0	*		*	*
1	*	*		
0	*	*		*
0	*	*	*	
0	*	*	*	*

[a] (= max(Counts) + 1000; adjusted from 393)

Constraints	Weights	p
*V(g)n	-6.78	< .0001
*i:	-3.52	< .0001
*VCvV	-1.54	< .0001
*e:	-0.23	< .0001
Ranking		**p**
*V(g)n >> {*i:, *VCvV, *e:}		.0002
*i: >> {*VCvV, *e:}		< .0001
*VCvV >> *e:		<.0001

(Note that the first ranking test gives the largest *p* value we've seen so far for the English lexical patterns, though it's still highly significant.)

Conclusions

- Phonological argumentation involves comparing type frequencies (class sizes)
- Small samples can suffice for statistical significance, under the right conditions
- OT ranking can be tested from type frequencies, even without direct interaction
- Some claimed interactions in English pass the test, while others are more doubtful

References

Alcántara, Jonathan. B. 1998. *The architecture of the English lexicon*. Cornell University PhD dissertation.

Borowsky, Toni. 1990. *Topics in the lexical phonology of English*. New York: Garland Publishing.

Chomsky, Noam, and Morris Halle. 1968. *The sound patterns of English*. Cambridge, MA: MIT Press.

CMU Pronouncing Dictionary. 1998. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26:289-337.

Halle, Morris, and K. P. Mohanan. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16 (1):57-116.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379-440.

Hempel, Car. G. 1945. Studies in the logic of confirmation (I). *Mind* 54 (213):1-26.

Kiparsky, Paul. 1976. Abstractness, opacity, and global rules. In Andreas Koutsoudas (ed.) *The application and ordering of phonological rules*, pp. 160-184. The Hague: Mouton.

Legendre, Géraldine, Antonella Sorace, and Paul Smolensky. 2006. The Optimality Theory - Harmonic Grammar connection. In Paul Smolensky and Géraldine Legendre (eds.) *The harmonic mind: From neural computation to Optimality-Theoretic grammar, vol. 2*, 339-402. Cambridge, MA: MIT Press.

McCarthy, John J. 2007. *Hidden generalizations: Phonological opacity in Optimality Theory*. London: Equinox.

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1):83-127.

Myers, James. 1993. A processing model of phonological rule application. PhD dissertation, University of Arizona.

Myers, James. 2008. Bridging the gap: MiniCorp analyses of Mandarin phonotactics. *WECOL* 37: 137-147.

Myers, James. Forthcoming. Automated collection and analysis of phonological data. In S. Winkler & S. Featherston (eds.) *Linguistic evidence*. Berlin: Mouton de Gruyter.

Pater, Joe. Forthcoming. The locus of exceptionality: Morpheme-specific phonology as constraint indexation. In Steve Parker (ed.), Phonological argumentation: Essays on evidence and motivation. London: Equinox.

Prince, Alan. 2007. Let the decimal system do it for you: A very simple utility function for OT. Rutgers University ms. ROA 943-1207.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in generative grammar*. Oxford, UK: Blackwell Publishing.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reiss, Charles. 2008. Constraining the learning path without constraints, or the OCP and NoBanana. In A. Nevins & B. Vaux, (eds.) *Rules, constraints and phonological phenomena*. Oxford: Oxford University Press.

Rubach, Jerzy. 1984. Segmental rules of English and cyclic phonology. *Language* 60:21-54.

Rubach, Jerzy. 1996. Shortening and ambisyllabicity in English. *Phonology* 13:197-237.

Shaw, Jason Anthony. 2007. Compensatory lengthening via mora preservation in OT-CC: Theory and predictions. *NELS* 38.