Adam Albright (MIT)
albright@mit.edu

UCLA Linguistics Colloquium
January 13, 2012

Additive markedness interactions in phonology

1 Markedness thresholds and cumulative constraint interactions

(1) A common intuition: rules/constraints evaluate strings independently

- Each rule/constraint has a structural description
  - #CC (initial clusters)
  - $\begin{bmatrix} -\text{son} \\ +\text{voi} \end{bmatrix}\begin{bmatrix} -\text{son} \\ +\text{voi} \end{bmatrix}$# (final clusters of voiced obstruents)
- Violation/application is triggered whenever structural description is met, regardless of what occurs elsewhere in the word
- Not observed, in English or elsewhere (?)
  - Final /bd/ repaired by epenthesis only if there are multiple voiced clusters in the word: [ɹʌbd] 'rubbed', but [əbzɔɹbəd] 'absorbed'
  - Final /bd/ repaired by epenthesis only if the word has an onset cluster: [ɹʌbd] 'rubbed', but [gɹæbəd] 'grabbed'
- If a structure is tolerated, it is tolerated regardless of violations elsewhere in the word
  - No 'markedness threshold' beyond which structures suddenly become intolerable

(2) A surprising pattern: Lakhota (Siouan)

- Roots may contain a wide variety of structures, including fricatives, aspirated/ejective obstruents, and consonant clusters
- However, unexpectedly few roots with combinations of these structures
  - Two clusters (***gleʃka*** 'spotted', ***wikt͡ʃemna*** '10')
  - Two fricatives (***ʃoʃa*** 'bubbly')
  - Cluster + aspirate/ejective combinations (***x'akpa*** 'curved/ruffled', ***blet͡ʃʰa*** 'decline in health')
  - Cluster + fricative combinations (***mnuɣa*** 'crunch on')
- Cumulative effect: multiple marked structures interact to make word less likely

(3) A similar phenomenon in English (other examples below)

- English allows initial #Cl clusters (***clip***, ***black***) and final *s*C# consonant clusters (***list***, ***lisp***, ***risk***)
- However, unexpectedly few words combining #Cl and sC#
  - Just two in CELEX: ***blast***, ***clasp*** (****glisp***, ****clesk***, ****ploast***, ****flisk***, …)
- Gradient restriction: morphemes can end in sC] unless there's a [Cl onset

(4) Claims of this talk

- Cumulative phonological restrictions: combinations of independently dispreferred structures are worse than expected
- A gradient (statistical) effect (not a categorical ban on combinations of otherwise perfectly acceptable structures)

- A static phonotactic effect (no alternations when violations are created through affixation)
  - Lakhota: ***x'an-kte*** 'work-FUT'
  - English: *glossed* [**glɔst**], *placed* [**pleɪst**] (*[glɔsəd], *[pleɪsəd])
- These effects, and the observed restrictions, can be derived with a model in which…
  - Candidates compete according to the sum of their weighted violations
  - A 'threshold' constraint imposes a markedness cutoff on acceptability of words to derive static phonotactic patterns

(5) Outline

- Demonstrating the effect: a lexical study of Lakhota
- The weighted constraint model: schematic illustration
- Applying the model: learning a grammar from the Lakhota lexicon
- Parallel example from English (with evidence about native speaker intuitions)
- Residual issues…

2 Cumulative markedness interactions: Lakhota

2.1 The basic pattern

(6) Background on Lakhota (Siouan)

- Stops and affricates: three-way contrast

		Initial		Medial	
Voiceless unaspirated	p, t, t͡ʃ, k	**k**a	'there'	ʃ'a**k**a	'strong'
Aspirated	pʰ, tʰ, t͡ʃʰ, kʰ	**kʰ**a	'to mean'	mã**kʰ**a	'earth'
Ejective	p', t', t͡ʃ', k'	**k'**a	'to dig'	t͡ʃi**k'**ala	'small'

- Fricatives: similar three-way contrast

		Initial		Medial	
Voiceless	s, ʃ, x, h	**x**ã	'scab'	i**x**a	'to laugh'
Voiced	z, ʒ, ɣ	**ɣ**ã	'messy hair'	ho**ɣ**ã	'fish'
Ejective/glottalized	s', ʃ', x'	**x'**ã	'to do'	ptu**x'**a	'to crumble'

- Sonorants: m, n, l, w, j

(7) Syllable structure

- Syllables generally of the form C(C)V
- Extensive set of CC onset clusters
  - Unaspirated stop + sonorant: *blo*[1] 'potato', *gmũ* 'twisted'
  - Unaspirated stop + obstruent: *kʃu* 'to bead', *ksapa* 'wise', *kte* 'to kill', *tka* 'heavy'
  - Fricative + C: *spã* 'to thaw', *xt͡ʃa* 'to bloom', *xmũ* 'to buzz', *ʃlo* 'to melt'
  - Nasal + nasal: *mni* 'water'

(8) Root shape

- Lakhota words can become quite long through affixation

wa-	*ju-*	*o-*	*ki-*	*wãʒi-la*	'unite'
indef.	obj.-causation-	loc.-	to/for-one-	dimin.	

[1] Unaspirated stops become voiced before sonorants in clusters (and a short epenthetic vowel is typically inserted: [bᵊlo]).

- Clusters, fricatives, ejectives, etc., are mostly found in roots, which are often quite short
  - C_0V, $C_0VC(V)$[2], C_0VC_0V
  - Longer monomorphemic roots: *paɣõta* 'duck', *wagmiza* 'corn', *it͡ʃamna* 'snow'
  - However, long roots frequently behave as if they are morphologically complex (e.g., other morphemes are 'infixed' in the middle of them, or other words with shared pieces and vaguely related meanings; for discussion, see Boas and Deloria 1941, pp. 26–28)
- Affixes (especially, prefixes) tend to use simple syllable structure, unmarked segments (sonorants, voiceless unaspirated stops, voiceless fricates)—e.g.,

Locative	*o-*, *i-*, *a-*, (*e-*), *ki-*	Causative	*-ja*, ***-kʰija***
Valence	*wa-*, *a-*, *ki-*	Tense	*-hã*, ***-s'e***, ***-kte***
Manner	*ka-*, *na-*, *pa-*, *ja-*, *ju-*		
Person markers	*wa*, *ja*, *ũ(k)*, *ma*, *ni*, ***t͡ʃʰi***; reflexive ***-it͡ʃ'i-***		

- Reduces the probability of encountering multiple clusters, ejectives, etc., in a word
  - Plausibly attributed to the morphological structure of words + phonological restrictions on affixes, not a ban on multiple clusters/fricatives/etc.

☞ Restrictions are most interesting within roots, where marked structures are freely allowed

(9) Within roots: statistically gradient cooccurrence restrictions

- Carter (1974, p. 51): general restriction of one cluster per morpheme
  - Many words like ***ks**apa* 'wise', *i**gm**u* 'cat'
  - Few words like ***gl**e**ʃk**a* 'spotted', *wi**kt͡ʃ**e**mn**a* '10'
- Informal observation suggests that roots with two fricatives are also scarce
  - Lots of words like *ka**ɣ**e* 'to make', ***s**apa* 'black', ***ɣ**opa* 'snore'
  - Few words like ***x**u**ɣ**a* 'dent', ***ʃ**o**ʃ**a* 'bubbly'
  - Virtually no words like ****ɣ**o**ʃ**a* (distinct fricatives)
- Combinations of fricatives with clusters are also rare
  - A few cluster-first examples: ***mn**u**ɣ**a* 'eat crunchily', ***ʃl**o**ɣ**a* 'make hominy'
  - Very few fricative-first examples: ***ɣ**o**pt**ã* 'listen', ***ʃ**a**kp**e* 'six'
- Roots with ejective + cluster combinations also rare
  - Few words like ***pt**u**x'**a* 'crumble'

2.2 A lexico-statistical study

(10) The Lakhota lexicon

- Constructed database of 12,396 entries in Buechel and Manhart (2002) dictionary
  - All entries except those beginning with *a-*, *i-*, *o-* (omitted since mostly prefixed, and could be left out without systematically undercounting any particular class of consonants)
- Rough morphological parsing to identify roots
  - Separated recurring roots and prefixes (distributional criterion)
  - Used meaning as a check to reduce hyperanalysis
    hoɣã 'fish' ≠ *ho*+*ɣã* 'bushy+voice'; *tʰaspã* 'apple' ≠ *tʰa*+*spã* 'thaw+ruminant's body part'

[2] There is a set of roots that are sometimes analyzed as /C_0VC/, but which always surface with a final vowel; see Shaw (1980) for discussion.

- Extracted mono- and di-syllabic morphemes
  - Removed reduplicated and contracted forms (show independent phonological restrictions)
  - Removed allomorphs derived by final vowel changes (‘ablaut’)
  - Removed morphemes longer than 2 syllables, as more likely to be complex or misanalyzed
  - Result: 2,275 entries (351 monosyllabic, 1924 disyllabic)
- Caveat: many entries that are almost certainly morphologically complex remain un-decomposed, and some words probably also over-decomposed
  - Less of a problem in 1–2 syllable morphemes, which we focus on here
- One other caveat: marking of aspiration in dictionary appears to be incomplete, especially for certain segments in initial position (e.g., [tʃ] vs. [tʃʰ])
  - I omit counts of aspirated stops in initial position

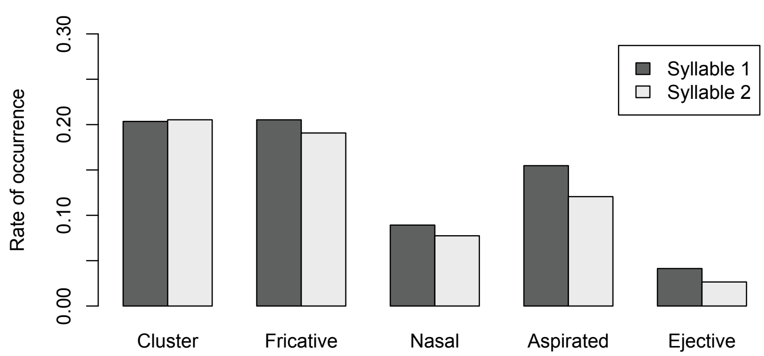
(11) Expected rates of occurrence for ‘doubly marked’ roots

- The question of interest: how often do we see roots that combine marked structures?
- Baseline (independence model):
  - Prob(C’VCCV) = Prob(initial C’) × Prob(medial CC)
- Are marked structures already expected to be rare in medial position due to positional restrictions?

(12) Positional effects within roots: very weak

- Clusters, fricatives, aspirated stops/affricates, and ejectives can all occur in either initial or medial position

	Initial	Medial
Clusters	***sk**uja* ‘sweet’, ***gl**epa* ‘vomit’	*hã**sk**a* ‘tall’, *hĩ**gl**a* ‘move suddenly’
Fricatives	***ʃ**apa* ‘dirty’, ***ʒ**ata* ‘forked’	*wa**ʃ**i* ‘work’, *wã**ʒ**i* ‘one’
Aspirates	***pʰ**ata* ‘to butcher’, ***kʰ**ãta* ‘plum’	*a**pʰ**a* ‘hit’, *wa**kʰ**ã* ‘holy’
Ejectives	***p’**o* ‘fog’, ***x’**oka* ‘sing with drum’	*nã**p’**ĩ* ‘wear around neck’, *i**x’**e* ‘stone’



- Nasals look rather infrequent when counted this way, but only 2 nasals, compared with 4 ejectives, 4 aspirates, 6 fricatives, and numerous clusters…
- Clusters, fricatives, nasals, aspirated/ejective stops are not appreciably rarer in medial position
- Thus, we expect that a reasonable number of disyllabic roots should have these structures in both positions simultaneously

(13) Characterizing the pattern informally: Observed/Expected counts

- Calculated rate of occurrence of consonant types in initial syllable (monosyllables and disyllables together) and second syllable (disyllables only)
  - E.g., clusters in first $\sigma = 20.4\%$ probability, fricatives in second $\sigma = 19.0\%$ probability
  - Joint probability = 3.9%
- Expect 75 CCVSV words in a lexicon of 1924 disyllables
- Only 45 CCVSV words actually occur
- O/E = 45/75 = .60 (somewhat underattested)

(14) Results

Initial	Medial							
	Cluster		Fricative		Aspirated		Ejective	
Cluster	43/80	? 0.54	45/75	? 0.60	11/47	*? 0.23	1/10	*? 0.10
Fricative	43/81	? 0.53	54/75	? 0.72	13/48	*? 0.27	10/10 =	✓0.96
Ejective	5/16	*? 0.31	13/15 =	✓0.86	0/10	* 0.00	8/2 =	3.80

(15) Restriction on cluster+cluster combinations (Carter 1974)

- Expect: 80 (20.4% in first syllable, 20.5% in second syllable)
- Actual: 43 (O/E = .54)
- 24 of the examples involve pseudoreduplication, and should perhaps be treated separately
  - *glegle*, *gligla*, *gmigma*, *gnãgnã*, *kpakpa*, *kpikpi*, *kpukpa*, *ktokta*, *kʃikʃa*, *kʃikʃã*, *psĩpsĩ*, *ptepte*, *pʃapʃa*, *skiska*, *stasta*, *stusta*, *xtʃʰaxtʃʰa*, *xpaxpa*, *xpixpi*, *xwaxwa*, *xwuxwu*, *ʃkaʃkã*, *ʃkiʃka*, *ʃtuʃta*,
- 19 distinct CC: *blaska*, *blaski*, *bluʃka*, *glamni*, *glapʃũ*, *gleʃka*, *gluʃtã*, *gnaʃka*, *ktʃʰãpta*, *kpamni*, *psexte*, *psextĩ*, *skokpa*, *skumna*, *skuxya*, *slitka*, *ʃkokpa*, *ʃkãgle*, *ʃtegla*
- Effect may be stronger than O/E value suggests, with numerous exceptions through reduplicative correspondence (Zuraw 2002)
- An OCP(cluster) effect?
  - Not easy to formulate, given standard assumptions about the kinds of structures that OCP constraints penalize (Suzuki 1998)
  - Also, as we are about to see, would be insufficient to capture range of restrictions
- Claim: a cumulative interaction between the individual clusters involved
  - Two violations in *gleʃka* worse than sum of individual violations in *glepa*, *ʃkata*
  - Such words are correspondingly rarer than expected

(16) Fricative-fricative combinations

- Expect: 75 (20.5% in first syllable, 19.0% in second syllable)
- Actual: 54 (O/E = .72)
- 44 of these involve exact or close copies (pseudoreduplicated)
  - *ʃaʃte*, *ʃĩkʃĩ*, *ʃkaʃkã*, *ʃkiʃka*, *ʃoʃa*, *ʃtuʃta*, *ʃuʃka*, *ɣĩɣe*, *ɣõɣa*, *ɣũɣa*, *kʃikʃa*, *kʃikʃã*, *pʃapʃa*, *psĩpsĩ*, *psũsũ*, *sasa*, *sãsã*, *sisa*, *sĩsĩ*, *s'ĩs'ĩ*, *skiska*, *stasta*, *stusta*, *sũsũ*, *xaxa*, *x'ax'a*, *xãxã*, *x'ãx'ã*, *xaxe*, *xaxla*, *xexe*, *x'ex'e*, *x'ex'e*, *xlaɣã*, *xloɣã*, *xloɣe*, *xloɣu*, *xmũɣa*, *xniɣa*, *xpaxpa*, *xpixpi*, *xuɣa*, *xwaxwa*, *xwuxwu*

- Only 10 items with truly distinct fricative-fricative combinations (and most involve a cluster)
  - *ʃix'ã*, *psexte*, *psextĩ*, *sekse*, *skuxya*, *ʃixtĩ*, *ʃĩxtĩ*, *ʃtaxe*, *xeʃma*, *xtʃãɣa*
- An OCP constraint on fricatives? (*[+cont]…[+cont])
  - Not a commonly observed OCP effect (cf. place, laryngeal features)
  - In fact, this would be the only clear-cut documented case that I'm aware of
  - Positing OCP([+cont]) predicts (wrongly?) languages in which fricatives become stops if there's a fricative earlier in the word
  - Positing OCP([+cont]) also predicts that fricatives specifically dislike coccurring with other fricatives; we'll see below that this isn't sufficient
- Claim: cumulative effect of two *Fricative violations

(17) Fricative-cluster combinations

- Expected: 156 (75 CCVSV, 81 SVCCV)
- Actual: 88 (45 CCVSV, 43 SVCCV) ⇒ O/E = .56
- Interaction of two distinct structures—not an OCP effect!
  - …and no particular reason to thing that there is a *CC…S or *S…CC constraint that specifically bans cluster+fricative combinations
- Can only be interpreted as cumulative interaction of constraints against clusters and fricatives

(18) Not all combinations are underattested

- Ejective-ejective combinations are overattested
  - However, nearly all pseudoreduplicated (*s'ĩs'ĩ*, *x'ãx'ã*, *t'ĩt'ĩ*, *kʰikʰi*, *tʰitʰi*, etc.)
  - Otherwise, near total ban on two aspirates/ejectives (MacEachern 1999; Gallagher 2010)
- Ejective-fricative combinations also overattested (will not say more about these here)

(19) A more rigorous test of the effect: a statistical model of the Lakhota lexicon

- Made a schematic version of the Lakhota lexicon, representing counts of 'major root shapes'
  - Took all 2,275 monosyllabic and disyllabic roots
  - Collapsed segments by place, manner: Stop (T), Fricative (S), Nasal (N), Lateral (L), Glide (J), h, Vowel (V)

    E.g., *ʃic'e*, *ʃok'ĩ*, *sĩc'i* ⇒ SVT'V; *hĩʃma*, *huxni* ⇒ hVSNV
  - Ignored aspiration on [tʃʰ] (contrast not marked systematically in dictionary, and aspirated [tʃʰ] appears not to pattern with other aspirates)
- For each of the 257 resulting root shapes, calculated number of occurrences in the lexicon

(20) Questions to test

- Can we predict how many times a given root shape occurs in the lexicon, based on the probability of the structures it contains (clusters, fricatives, aspirates, etc.)
- Do predictions improve if we allow additional penalties for combinations of marked structures?
- Approach: nested log-linear models, predicting rates of attestation of different root shapes based on the number of violations of different constraints
  - Start with 'simple' constraints (*Fricative, *Aspirate, *Ejective, *Nasal, various *CC)
  - Then, add constraints on combinations to see whether they offer significant improvement (likelihood ratio test)

(21) Baseline model: different classes have different frequencies

Constraint	Coefficient	Sig. (χ^2)
*Ejective	-1.434	<.0001
*Lateral	-0.941	<.0001
*Nasal	-0.821	<.0001
*Aspirate	-0.693	.012
*Fricative	-0.546	<.0001
*Glide	-0.191	<.0001
*Stop	0.459	<.0001

Constraint	Coefficient	Sig. (χ^2)
*CC:stop+stop	-1.634	<.0001
*CC:stop+nasal	-1.233	<.0001
*CC:stop+liquid/glide	-1.249	<.0001
*CC:fric+stop	-0.421	<.0001
*CC:fric+liquid/glide	-0.392	<.01
*CC:fric+nasal	-0.098	.511

- Negative coefficient means fewer roots containing the structure in the lexicon

(22) Doubly marked roots, part 1: two of the same structure

- Added constraints for *CC…CC, *S…S, *C^h…C^h, *C'…C', *N…N
- Likelihood ratio test shows significant improvement over baseline model (p<.0001)
- Specific effects

Constraint	Coefficient	Sig. (χ^2)
*CC…CC	-0.852	<.0001
*S…S	-0.517	<.0001
*N…N	-0.058	0.83

Constraint	Coefficient	Sig. (χ^2)
*C'…C'	0.433	0.24
*C^h…C^h	0.608	<.05

- Combinations of two clusters, two fricatives underattested (with stronger effect for clusters)
- Nasals cooccur as often as expected (no additional effect)
- *C'…C' and *C^h…C^h are surprisingly *overattested*
  - As noted above, virtually all pseudoreduplicated

(23) Doubly marked structures, part 2: distinct structures

↓C1/C2→	*CC	*Fric	*Asp
*Fric	-0.401 ***		
*Asp	-0.546 ***	-0.830 ***	
*Ejec	-0.980	*0.344* **	-1.241 **

- Confirms cline of underattestation observed informally above with O/E counts

(24) Summary so far: combinations of dispreferred structures are even rarer than expected

- Degree of underattestation is correlated with frequency of the parts
- Clusters and fricatives show moderate restrictions
  - Lowish frequency in Lakhota in spite of large inventories
  - Also cooccur with themselves and with each other somewhat less often than expected
- Aspirates and ejectives show stronger restrictions
  - Clusters and fricatives cooccur with aspirates rather rarely
  - Clusters cooccur with ejectives even more rarely
- Strongest effect: cooccurrence of ejectives with aspirates
  - May also be due to a more general ban on laryngeal cooccurrence

Important: underattestation **goes beyond what is expected** based on joint probability of the parts!

(25) Maybe we have the wrong expectations?

- Simple expectation: probability of cooccurrence AB = joint probability (Prob(A)×Prob(B))
- Frisch (1996, pp. 152–153): combinations of high frequency elements are likely to be attested, and support learning (or creation?) of additional words containing those combinations
- Combinations of low frequency elements are low frequency, and may even be accidentally unattested, so 'never get their foot in the door' (less support for creating)
- Rich get richer and poor get poorer effect → warping such that low expected frequency leads to even lower attested frequency. Perhaps that is all we're seeing here?

(26) Comparison: cooccurrence with nasals

- Nasals are relatively infrequent in initial position (≈9%)
- Combinations involving nasals are correspondingly rare, but not unexpectedly so

Initial	Medial			
	Cluster	Fricative	Aspirated	Ejective
Nasal	27/35 = ✓0.77	39/33 = ✓1.19	20/21 = ✓0.97	8/5 = ✓1.76

- Loglinear model: interactions with *Nasal are non-significant, with mostly positive coefficients (combinations slightly overattested)
- Conclusion: cumulative rarity alone is not enough to cause disproportionate underattestation

(27) Claim: a grammatical effect

- Although nasals are somewhat low in type frequency in Lakhota, they are not subject to all that many phonological restrictions
  - Can occur freely in roots, prefixes, and suffixes
  - Low type freq. due to small inventory of nasal place contrasts, rather than dislike of nasals
- The insight that this leads to: degree of underattestation of a combination follows from strength of **grammatical** constraints against the elements involved, not low frequency per se

(28) Summarizing this result

- Cumulative effects: dispreferred elements cooccur even less often than expected, based on degree of dispreference for either one independently
- Leads to appearance of 'complexity threshold': marked structures may occur individually, but combinations are restricted
- Crucially, not as simple as saying that combinations of rare things are extra-rare
  - Nasals are relatively infrequent, but not grammatically restricted; no cumulative effects
- We need a grammatical analysis that can distinguish markedness (=strength of grammatical penalty) from frequency (=observed probability)

2.3 Capturing cumulative interactions with an additive constraint model

(29) The challenge

- Allow some combinations of marked structures to occur as often as expected based on independent probability, while other combinations receive an additional penalty
- Degree of cumulative effect should follow from grammatical status of subparts
  - Combinations of things which are already marginal become even worse

- Effect is lexically gradient—perhaps universally so?
- A root/morpheme structure effect: no alternations
- In this section, we generate a lexicon with these properties, using a grammar of weighted constraints (Harmonic Grammar: Legendre, Miyata, and Smolensky 1990; Pater 2010)

(30) Such interactions cannot be derived in OT

- CC clusters tolerated: DEP(ə) ≫ *[CC

/kpã/ 'fine'	DEP(ə)	*[CC
☞ a. kpã		*
b. kəpã	*!	

- Fricatives tolerated: IDENT([±cont]) ≫ *Fric

/maza/ 'metal'	IDENT([±cont])	*Fric
☞ a. maza		*
b. mata	*!	

- Both automatically tolerated simultaneously

/kpaza/ 'dark'	DEP(ə)	IDENT([±cont])	*[CC	*Fric
☞ a. kpaza			*	*
b. kəpaza	*!			*
c. kpata		*!	*	
d. kəpata	*(!)	*(!)		

- Presence of one violation can't 'strengthen' severity of the other
- Put differently: *[CC and *Fricative violations can't "gang up" to require a repair

(31) Conjoined constraints are not the answer

- Narrowly speaking, it would be possible to capture the observed cooccurence restrictions with conjoined constraints

/kpaza/ 'dark'	*[CC & *Fric	DEP(ə)	IDENT([±cont])	*[CC	*Fric
a. kpaza	*!			*	*
☞ b. kəpaza		*			*
☞ c. kpata			*	*	
d. kəpata		*(!)	*(!)		

- However, this would fail to capture several key aspects of the pattern
  - Conjoined constraint is an independent parameter—no necessary relation between ranking of *X&*Y and subparts *X, *Y
  - Could produce categorical phonotactic restrictions as well as gradient ones
  - Could produce alternations in complex words
- Conjoined constraints are too powerful a mechanism for deriving cumulative effects
  - We need something that allows the penalty of two violations to depend on the penalty of the components

(32) Weighted constraint models: promising, but not an automatic solution

- Harmonic Grammar (Legendre, Miyata, and Smolensky 1990; Pater 2010)
  - Constraints are assigned numerical weights
  - For each candidate, constraint violations are multiplied by weights, and then summed
  - Promising: sum of multiple violations > individual violations

- Pater (2010): Markedness and Faith violations trade off, precluding many cumulative effects

/kpaza/ 'dark' Weight:		DEP(ə) 3	IDENT([±cont]) 3	*[CC 2	*Fric 2	Sum
☞ a.	kpaza			-1	-1	-4
b.	kəpaza	-1			-1	-5
c.	kpata		-1	-1		-5
d.	kəpata	-1	-1			-6

  - Repairing multiple $\mathcal{M}$ violations often requires an equivalent number of $\mathcal{F}$ violations
  - $w(\mathcal{F}) > w(\mathcal{M}) \rightarrow 2w(\mathcal{F}) > 2w(\mathcal{M})$
- If we want multiple markedness violations to gang up, we need to sidestep this trade-off between Markedness and Faithfulness

(33) Redefining the competition: the null parse

- Main task for producing a lexicon: determine well-formedness of strings
  - [blɪk] is well-formed, [bwɪk] is marginal, [bnɪk] is illegal
- Unfaithful mappings (converting ill-formed strings to well-formed ones) are needed elsewhere
  - Repair structures created by morpheme concatenation (*[kɪk-d] → [kɪkt])
  - Loanword adaptation (Dnieper → [nipɹ̩], [dənipɹ̩])
  - Let's set these aside for now, and focus on producing a lexicon that obeys static restrictions[3]
- Elimination of possible words: losing to the null parse (Prince and Smolensky 2004)
  - Null candidate represents no output ('gap')
  - Escapes all markedness violations, except a single violation of the constraint MPARSE
- Schematically:
  - Fricatives allowed: Weight(MParse) > Weight(*Fricative)

/ɣã/ Weight:		MPARSE 4	*[CC 3	*FRIC 3	Total
☞ a.	ɣã		-1		-3
b.	*null*	-1			-4

  - Clusters allowed: Weight(MParse) > Weight(*[CC)

/gla/		MPARSE:4	*[CC:3	*FRIC:3	Total
☞ a.	gla		-1		-3
b.	*null*	-1			-4

  - Combinations not allowed: Weight(*[CC) + Weight(*Fricative) > Weight(MParse)

/ɣãgla/		MPARSE:4	*[CC:3	*FRIC:3	Total
a.	ɣãgla		-1	-1	-6
☞ b.	*null*	-1			-4

- Two issues
  - Need the effect to be gradient, not categorical
  - Need to ensure that only those structures which are independently marginal can lead to cumulative effects

[3] For an approach that allows cumulative effects with standard faithfulness constraints, see Albright, Magri, and Michaels (2008)

(34) Gradient well-formedness

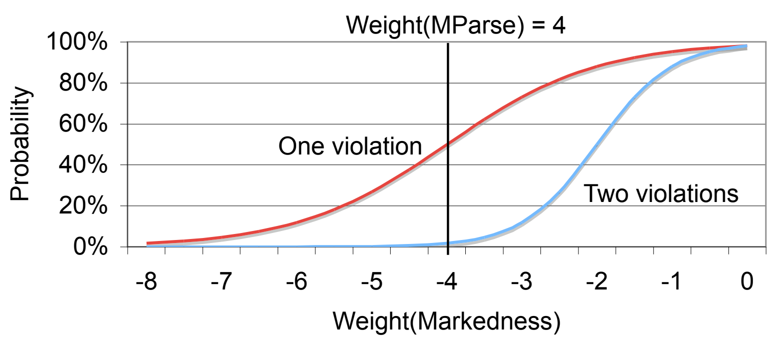
- The tableaux above simply assume that the candidate with the smallest penalty wins
- Following Maximum Entropy approaches (Goldwater and Johnson 2003; Jäger 2007):

$$\text{Prob(output)} = \frac{e^{\text{penalty(output)}}}{\sum_{\text{candidates}} e^{\text{penalty(candidate)}}}$$

  - ○ Penalty = weighted sum of violations
- For the schematic examples above:
  - ○ Prob(ɣã), Prob(gla) ≈ 73%
  - ○ Prob(ɣãgla) ≈ 12%

(35) The poor get poorer

- Under the scheme in (34), the farther the weight of a markedness constraint is below MParse, the more likely the structure is to surface
- However, this condition also makes it relatively unlikely that the markedness constraint can gang up to overcome MParse (e.g., 2×Weight(Markedness) > Weight(MParse))
- Gang effects occur in a relatively small band around the weight of MParse



  - ○ Discrepancy between probability of one vs. two violations is most noticeable when Weight(Markedness) is close to MParse
  - ○ Probability of single violation is also beginning to decline at this point
  - ○ Captures observation that cumulative effects are seen only for structures that are already penalized individually by the grammar

(36) Local summary

- Grammar of weighted constraints penalizes outputs with multiple violations more severely than outputs with single violations
- Weight of threshold constraint (MPARSE) establishes (gradient) cut-off
  - ○ Grammar probabilistically eliminates inputs, based on severity of markedness violations
  - ○ Eliminates = selects the null parse instead (i.e., produces no overt output)
  - ○ The fact that the null parse violates just one constraint (MParse) allows multiple markedness violations to gang up and produce cumulative effects
- Strength of cumulative effect depends directly on strength of simpler effects
  - ○ Summed violations of simpler constraints
  - ○ Cumulative effect not learned separately; follows from weights of individual constraints
  - ○ Precludes large disparities between well-formedness of single vs. multiple violations
  - ○ Predicts that cumulative effects should be primarily or exclusively gradient

2.4 Using this mechanism to produce the Lakhota pattern

(37) Deriving cumulative interactions with MParse

- The loglinear model at the end of section 2.1 is formally equivalent to a MaxEnt grammar, using the relevant factors as constraints
- In that section, we tested for cumulative effects by including factors for interactions
  - *CC…CC, *CC…Fric/Fric…CC, etc.
- In this section, we derive such effects with a simpler model, employing MParse

(38) Input to the model

- 257 schematic root types described above, with their type frequencies in the Lakhota lexicon
- Constraints:
  - Markedness constraints on simple structures (fricatives, aspirates, ejectives, nasals, various cluster types, etc.)
  - MParse
- Each attested root type competes with a "null" candidate

/TLVSV/		Freq	MParse	*[CC	*Fric	etc…
a.	TLVSV	10		-1	-1	
b.	*null*		-1			

- Initial bias: weight(MParse) $>$ weight($\mathcal{M}$)
  - Model is biased not to produce structures unless there is positive evidence for them[4]
  - Weight($\mathcal{M}$) $=10$
  - Weight(MParse) $=5$
- Learning: MaxEnt Grammar Tool (Hayes, Wilson, and George 2009)

(39) The resulting grammar

Constraint		Weight
MParse	Penalizes null output	16.34
*TN	*Stop + nasal	8.86
*SN	*Fricative + nasal	8.85
*Ejective		8.62
*SR	*Fricative + liquid/glide	8.49
*TT	*Stop + stop	8.44
*ST	*Fricative + stop	8.24
*Asp	*Aspirated [−strid] stop	7.63
*TR	*Stop + liquid/glide	7.13
*SV	*Prevocalic fricative (McAllister 2009)	6.95
*Glide		6.06
*Lateral		5.98
*Nasal		5.25
*CC		0.00
*Stop		0.00

[4]Technically, this is implemented by assigning mean "target" values for weights (μ), as shown above, plus standard distributions that determine how strictly the model is constrained to find weights around those values ($\sigma^2 = .1$).

- Weight of MPARSE violation is substantially above weights of other violations, independently
- However, rarer structures have weights > .5×weight(MPARSE)
- This puts multiple violations within 'striking distance' to overcome MPARSE

(40) Predictions of the grammar: probability of output (representative selection)

Simple		With self		Other combinations			
NV	100%	NVNV	100%	NVT^{h}V	97%	NVSV	98%
SV	100%	SVSV	92%	SVT^{h}V	85%	SVTTV	72%
T^{h}V	100%	T^{h}VT^{h}V	75%	T^{h}VSV	85%	T^{h}VTTV	43%
TTV	100%	TTVTTV	37%	TTVSV	72%	TTVT^{h}V	43%

- All simple structures assigned 100% probability
- Combinations with nasals not significantly penalized
- Other combinations penalized to varying extents (clusters > aspirates, fricatives > nasals)
  - Exact probability depends on parameters of the model; have not attempted here to match lexical probabilities precisely

(41) Summary of this section

- Schematic illustration of how a grammar of weighted constraints, including MPARSE, can yield cumulative effects
- Strength of the effect depends on weight of constraints on simple structures
- Qualitative match to observed effect: combinations of dispreferred structures are even more strongly underattested

3 English cluster combinations

3.1 Cumulative effects in the English lexicon

English also shows (gradient) cumulative effects

(42) A widely discussed effect: clusters + OCP

- OCP violations are underattested, but legal in English (Berkley 2000)
  - *pipe, beep, bib, mime, kick, cake, gag, gang…*
  - *lull, loll, rare, roar*
- Non-coronal #sC clusters are also somewhat underattested, but fine
  - *spy, spike, sky, skate*
- Combinations of the two are rare or missing
  - *spVp, *skVk, *slVl (also: *plVl) (Browne 1981; Davis 1984; Frisch 1996; Coetzee 2008)
- Frisch (1996): such combinations are rarer than expected, given independent frequency of clusters and OCP violations

(43) Another (previously unnoticed?) example: CR…*s*C

- English has words that begin with obstruent+liquid clusters
  - pr, pl, br, bl, fr, fl, etc.

- English has words that end with *s*+stop clusters
  - sp, st, sk
- Surprisingly few words with Cr…sC combinations
  - *clasp*, *trust*, *grist*
  - Many combinations completely lacking, esp. with *Cl*-: **pl*V*sk*, **pl*V*st*, **bl*V*sk*, **gl*V*sp*, **gl*V*st*

(44) The general claim

- Certain combinations, such as CR…sC, are unexpectedly rare in English, relative to frequency of subparts
- However, not sufficient to simply point out that there are no words like *[plɪsk], *[drʌsk], etc.
- As with Lakhota, we must show that there are fewer than expected
- Compare: no words like *[sfælk], *[sfɪlt], but this is unsurprising, since #sf words are rare to begin with

(45) Testing this effect with a loglinear model

- Took all attested onsets and codas in monosyllabic lemmas listed in CELEX (Baayen, Piepenbrock, and van Rijn 1993)
  - Final r-clusters not present in CELEX transcription; recovered from orthography
- For each logically possible combination (n=7081), took type count of lemmas
  - Some combinations quite frequent: 28 fVl, 25 rVt, 23 bVt, etc.)
  - Most (n=5565) are unattested: zVrv, nVsp, grVsk
  - Some gaps are principled (*skVk, *klVl, etc.), but many (most?) are arguably accidental
- As above, goal of model is to predict relative likelihood of different combinations

(46) The baseline model: constraints on individual structures

- Coded ‘coarsely’ in order to avoid overfitting
  - Heuristic: categories that define categorical restrictions in English cluster types
  - P,B = stops, F = fricative, N = nasal
- Segment identity: positional (onset, coda)
  - Stop, affricate, fricative, voiced fricative, nasal, *l*, *r*, *s*
- Syllable structure:
  - Onset, Coda, Complex Onset, Complex Coda
- Cluster types
  - Onset: Pr, Br, Fr, Pl, Bl, Fl, sp, st, sk, sN
  - Coda: lP, lB, lF, lN, rP, rB, rF, rN, sp, st, sk, NP, NB, NF, PP, PS
  - Coda obstruent+obstruent, voiced obstruent+obstruent
- Baseline model: predicts some differences in rates of attestation, based on frequencies of parts

(47) A relevant phonotactic restriction: OCP

- Many cluster-cluster combinations incur OCP violations
  - *sm*V*mp*, *sp*V*lp*, *pr*V*sp*, etc.
  - As discussed, these are well known to affect rate of attestation

- Added factors for simple OCP violations
  - Place: OCP(lab), OCP(dors), OCP(cor)
  - Manner: lVl, rVr, sVs
- Result: small but significant improvement ($p(>|\chi|) < .0001$)

(48) Cumulative effects: *CR…CC

- Added factors for combinations of clusters
  - ClVsC, CrVsC, ClVNC, CRVNC
  - Question: are words like *bl*V*sk*, *sk*V*lt* significantly underattested?
- Result: significant improvement ($p(>|\chi|) < .01$)
- However, effect is seen only for some combinations
  - ClVsC significantly underattested ($p(>|\chi|) < .01$), but CrVsC not ($p(>|\chi|) = .9$)
  - Combinations with nasal clusters are *overattested* (ClVNC: $p(>|\chi|) <.005$; CrVNC n.s.)

(49) Summary of this section

- Gradient cumulative effects also seen in the English lexicon
- As with Lakhota, some combinations strongly underattested, while others occur as often (or even more often) than expected
  - #Cl clusters, but not #Cr clusters
  - sC# clusters, but not NC# clusters
- General line of attack:
  - Derive penalty for combinations from penalty on simple structures
  - For this to work: grammar of English must prefer #Cr, NC# over #Cl, sC#
- On-going project: does this preference fall out from the lexical statistics of English, or does it reflect a phonological preference?

3.2 Cumulative effects in acceptability judgments

(50) Testing for speaker knowledge of cumulative effects

- Strategy: make up nonce words with one or two clusters

Onset CC	brVC, spVC, dwVC, …
Coda CC	CVmp, CVnt, CVsp, …
Both	brVmp, blVnt, dwVsp, …

- Ratings of brVC, brVCC, etc. allow us to estimate the weight of *br
- Ratings of CVndʒ, CCVndʒ, etc. allow us to estimate the weight of *ndʒ
- Question: do these values suffice to predict the rating of brVndʒ?
  - Or, is the rating of brVndʒ even lower than expected based on its parts?

(51) A nonce-word acceptability experiment

44	Onset CC only	pl, kl, bl, gl, pr, kr, br, gr, kw, θw	bɹɛn, blʌs, klædʒ, dwɛp, …
26	Coda CC only	mp, nt, ndʒ, sp, st, sk, lk	dɛmp, lænt, ɹɛndʒ, kɛst, …
41	Onset + Coda CC	Combinations (non-identical place)	bɹæsk, blɛnt, klɛsp, dwɪmp, …

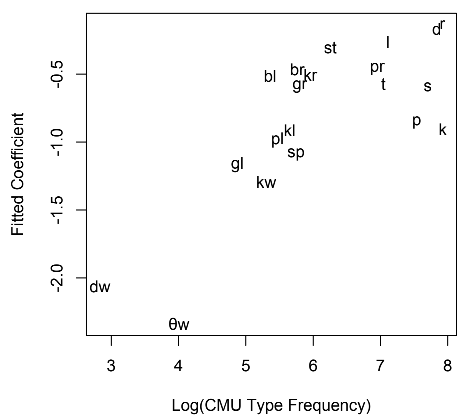
- Also 78 fillers (no clusters, non-English clusters)

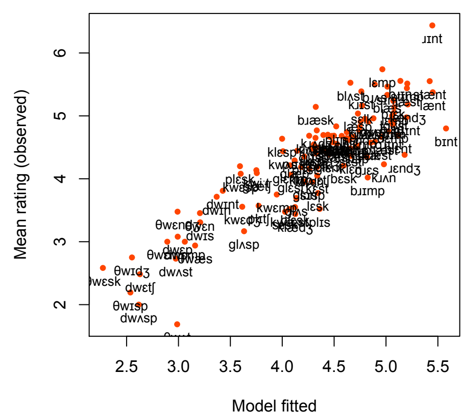
(52) A web-based experiment

- Read in carrier sentences by two linguistically trained native speakers of American English
- Subjects heard each novel word used as either a noun or verb (randomized and counterbalanced across subjects)
- Task: rate "how plausible you think it sounds as a potential word of English" from 1 (worst) to 7 (best)
- Presented using Experigen (Becker and Levine 2010)
- 91 subjects recruited with Amazon Mechanical Turk (limited to US, self-reported native speakers)

(53) Step 1: modeling ratings with "simple" phonological factors

- Fitted linear mixed-effects model (using `lme4` in R), predicting ratings based on:
  - ○ Identity of the vowel
  - ○ Identity of the onset (or onset cluster)
  - ○ Identity of the coda (or coda cluster)
  - ○ Random effect for subjects
- Two reassuring observations
  - ○ Preference for/against a structure tends to correlate with frequency (shown here for onsets)
  - ○ Acceptability ratings of words pretty well predicted by phonological factors





(54) Step 2: Cumulative effects

- Added factors penalizing combinations of clusters, coded broadly
  - ○ ClVNC, CrVNC, CwVNC, ClVsC, CrVsC, CwVsC, sCVlC, sCVNC
- Nested model comparison: do these factors offer significant improvement over model in (53)?
- Yes: $\chi^2(8) = 35.91$, $p < .0001$
- The goodness of a word like *dwesp* cannot be fully predicted based on its subparts alone
  - ○ Need an additional penalty for combinations

(55) Which combinations are penalized?

- Claim above: combinations of independently bad structures are even worse than expected
- Comparison

	Coda	✓NC	? sC
Onset	✓bl	blVNC	blVsC
	? dw	dwVNC	dwVsC

- Coding above is slightly too coarse to test this, since classes “Cr”, “Cl”, “Cw” are not homogenous groups (e.g., kw ≻ dw ≻ θw)
- Refitted model, with just factors penalizing specific cluster combinations:

dw.NC
β = -0.241 (n.s.)

bl.sC
β = -0.321 (n.s.)

→ dw.sC ←
β = -0.722 ($p<.05$)

(*Effect for dwVsC is stronger, and significant*)

4 Summary and discussion

(56) Cumulative interactions of markedness constraints

- Doubly marked structures less good than expected, given independent acceptability of subparts
- Seen in lexical counts, acceptability ratings
- Not discussed here: also seen in child simplifications (Levelt, Schiller, and Levelt 2000; Albright, Magri, and Michaels 2008)
- An underdocumented, but perhaps quite prevalent effect

(57) Analytical challenges

- Explain why markedness violations can gang up to produce static phonotactic effects, but evidently do not lead to repairs or alternations
- Explain why cumulative effects are seen for some combinations, but not others

(58) Pieces of the account

- Grammar of weighted constraints
  - Multiple violations may sum across the word
  - Creates potential to make multiple violations worse than single violations
- Competition with null candidate derives lexical underattestation
  - Weight of MPARSE constraint establishes ‘complexity threshold’
- An intrinsically gradient effect
  - Nature of competition in MaxEnt models means that doubly marked structures will be penalized only if singly marked structures are already somewhat degraded
  - Strength of effect depends on status of independent subparts

(59) An immediate question: more dramatic effects for triply-marked structures (and beyond)?

- A perennial worry about additive models: elaborate counting effects
  - "Morphemes may have three but not four fricatives"
  - "Morphemes may have at most two instances of: fricatives, aspirates, or clusters"
- The proposed model does indeed predict such effects—e.g., Lakhota:
  - $4\times$Weight(*Nasal) $>$ Weight(MParse) $>$ $3\times$Weight(*Nasal)
  - Weight(*Nasal) $+$ Weight(*Lateral) $+$ Weight(*Aspirate) $>$ Weight(MParse)
- It's conceivable that such effects really do occur
  - Hard to evaluate based on lexical counts, because expected frequencies of such words are already very low
  - Intuitively, hypothetical words can become quite unacceptable when they involve multiple independently OK structures: English [səˈflʊk], [ˈtækɹəp]

(60) The relation between static phonotactics and alternations

- The mechanism proposed here cannot derive alternations
  - Pater (2009): once faithfulness is involved, violations 'trade off' in a way that precludes different repairs depending on number of markedness violations
- This seems good: evidently no alternations triggered by cumulative complexity thresholds
  - …and they probably would have been noticed, since alternations have been examined more comprehensively than static phonotactics[5]
- But why don't we simply get paradigm gaps when affixed words reach the threshold?
- One possibility: different levels of evaluation
  - Root-level evaluation of 'Morpheme Structure Constraints' vs. stem/word-level evaluation of morphologically composed forms
- Another possibility: threshold increases with morphological complexity
  - Null output for bimorphemic word = two 'MParse' violations

(61) Future work

- Further experimental work: to what extent do speakers encode such effects?
  - Lexicon is gappy, hard to test predictions about combinations already predicted to be rare
  - Results above indicate that speakers do dislike words like *drusk* more than expected based on subparts
  - Need broader confirmation, with a wider variety of combinations
- What kinds of violations can interact?
  - Examples tested here involve 'similar' structures (consonants; onset, coda clusters)
  - Model is quite general, and predicts broader range of effects

[5] Though for a possible example, see Green (2010)

References

Albright, A., G. Magri, and J. Michaels (2008). Modeling doubly marked lags with a split additive model. In H. Chan, H. Jacob, and E. Kapia (Eds.), *BUCLD 32: Proceedings of the 32nd Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Baayen, R. H., R. Piepenbrock, and H. van Rijn (1993). *The CELEX lexical data base on CD-ROM*. Philadelphia, PA: Linguistic Data Consortium.

Becker, M. and J. Levine (2010). Experigen: An online experiment platform. Available at https://github.com/tlozoot/experigen.

Berkley, D. M. (2000). *Gradient OCP Effects*. Ph. D. thesis, Northwestern University.

Boas, F. and E. Deloria (1941). *Dakota Grammar*, Volume 23 of *Memoirs of the National Academy of Sciences*. Washington: United States Government Printing Office.

Browne, W. (1981). Slavic -ba and english *slil: Two persistent constraints. *Folia Slavica 4*, 219–226.

Buechel, E. and P. Manhart (2002). *Lakota dictionary: Lakota-English/English-Lakota*. Lincoln, NE: University of Nebraska Press.

Carter, Jr., R. T. (1974). *Teton Dakota Phonology*. Ph. D. thesis, University of New Mexico.

Coetzee, A. (2008). Grammaticality and ungrammaticality in phonology. *Language 84*, 218–257.

Davis, S. M. (1984). Some implications of onset-coda constraints for syllable phonology. In *Chicago Linguistic Society*, Volume 20, pp. 46–51.

Frisch, S. (1996). *Similarity and Frequency in Phonology*. Ph. D. thesis, Northwestern University.

Gallagher, G. (2010). Perceptual distinctness and long-distance laryngeal restrictions. *Phonology 27*, 435–480.

Goldwater, S. and M. Johnson (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson, and O. Dahl (Eds.), *Proceedings of the Workshop on Variation within Optimality Theory, Stockholm University*, pp. 111–120. Stockholm University.

Green, C. (2010). *Prosodic phonology in Bamana (Bambara): Syllable complexity, metrical structure, and tone*. Ph. D. thesis, Indiana University.

Hayes, B., C. Wilson, and B. George (2009). Maxent grammar tool. Java program.

Jäger, G. (2007). Maximum entropy models and Stochastic Optimality Theory. In A. Zaenen, J. Simpson, T. H. King, J. Grimshaw, J. Maling, and C. Manning (Eds.), *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*, pp. 467–479. Stanford: CSLI Publications.

Legendre, G., Y. Miyata, and P. Smolensky (1990). Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. Technical Report 90-5, Institute of Cognitive Science, Univ. of Colorado.

Levelt, C., N. O. Schiller, and W. J. Levelt (2000). The acquisition of syllable types. *Language Acquisition 8*, 237–264.

MacEachern, M. R. (1999). *Laryngeal Cooccurrence Restrictions*. Garland Publishing.

McAllister, T. (2009). *The Articulatory Basis of Positional Asymmetries in Phonological Acquisition*. Ph. D. thesis, MIT.

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science 33*, 999–1035.

Prince, A. and P. Smolensky (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing.

Shaw, P. A. (1980). *Theoretical Issues in Dakota Phonology and Morphology*. New York: Garland Publishing.

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

Zuraw, K. (2002). Aggressive reduplication. *Phonology 19*, 395–439.