

Class 10, 5/2/2018: Paradigm Uniformity I

1. Assignments

- Read: (2015) Bruce Hayes and James White, Saltation and the P-map. *Phonology* 32:267-302.
 - Purposes: case of diachronic explanation, getting ready for experimental studies of paradigm uniformity next time.
- Come talk with me about your term paper if you haven't already.

2. Today

- Finish a tiny bit about acquisition: hidden structure
- Start in on Paradigm Uniformity, mostly with a Socratic data-wallow.

3. Comments on homework

- Nice work, I enjoyed reading.
- Bits that intrigued me:
 - Modeling frequency per Appendix (a challenge, since some pronunciations are Mennian relics, but worthwhile)
 - Greater violation in long words — is there a baseline *STRUC constraint, ganging with other Markedness?
 - Weight drift — whereas *languages* should not be expected to have coherent weight-drift across time (due to discontinuity of acquisition), perhaps individual children should? Or are there discontinuities?
 - Greater Faithfulness for proper names (any precedent for this? Names do have particular phonology in many languages.)

A LITTLE BIT ON HIDDEN STRUCTURE

4. Definition and examples

- Aspects of representations not inferable from surface form
- Examples:
 - underlying representation (German [rat] = /rad/, /rat/)
 - metrical feet (two ways to bracket a trisyllables with penultimate stress)
 - syllabification ([ab.ra] vs. [a.bra], with consequences for stress, metrics)

5. Why is hidden structure hard to learn?

- If you make an assumption about feet, then all the rest of the grammar must be tailored to that assumption.
- But most ranking/weight algorithms blindly try to optimize all the constraints at once.

6. A toy example: mini-German

- Example drawn from:
 - Pater, Joe, Robert Staubs, Karen Jesney and Brian Smith (2012) Learning probabilities over underlying representations. In the *Proceedings of the Twelfth Meeting of the ACL-SIGMORPHON: Computational Research in Phonetics, Phonology, and Morphology*. 62-71.
- We need only consider four data:
 - ‘advice-plain’ /rat/ → [rat]
 - ‘advice-suffixed’ /rat-a/ → [rata]
 - ‘wheel-plain’ /rad/ → [rat]
 - ‘wheel-suffixed’ /rad-a/ → [rada]
 - N.B. *-a* is not a suffix in German but it is easy to type.
- For learning, let’s explore the larger set of candidates that arises if we are trying to learn UR’s.
 - No particular reason to think ‘advice’ is anything other than /rat/.¹
 - But ‘wheel’ has two candidates, /rat/, /rad/.

advice-plain	/rat/	→	☞ [rat]
advice-suffixed	/rat-a/	→	☞ [rata]
wheel-plain	/rad/	→	☞ [rat]
	/rad/	→	[rad]
	/rat/	→	☞ [rat]
	/rat/	→	[rad]
wheel-suffixed	/rad-a/	→	☞ [rada]
	/rad-a/	→	[rata]
	/rat-a/	→	☞ [rada]
	/rat-a/	→	[rata]

7. What defines success?

- We must derive *at least one* of the observed ☞ candidates for each input.
- We must impose **consistency** on the UR’s, since we need a good UR to pass a wug test on future forms.

¹ Actually, people occasionally override the “what you see is what you get” principle for non-alternating morphemes when they do “set up as”: set up all [h] as /x/, so it can trigger velar place assimilation (Toba Batak); then revert all /x/ to [h] on the surface. This is not so commonly done as it used to be ...

- I believe that it will not do, as Pater et al. suggest, to let the UR vary freely within its own paradigm.

8. Constraints

- Let's not bother with constraints that would derive Intervocalic Voicing, since /rat-a/ → [rata] will straightforwardly remove this possibility.
- We do need the standard constraints for Final Devoicing:
 - * $\begin{bmatrix} \text{-sonorant} \\ \text{+voice} \end{bmatrix}$]word
 - IDENT(voice)
- We need, following Boersma, Appousidou, Pater et al., constraints that force a particular allomorph as the UR.
 - WHEEL IS /rad/ — correct!
 - WHEEL IS /rat/ — wrong!
 - Apoussidou, D. (2007). *The learnability of metrical phonology*. Utrecht: LOT

9. A fancier kind of tableau: collating over hidden structures

- Observed candidates sum over all their sources.
- You win if the frequency of the correct observed candidate is 1.
- The weights were established by me, using thought.
 - Socrates: justify them, remembering that this is maxent.

	Hidden	Overt freq	wheel /rad/	wheel /rat/	*Coda Voiced Obs	Ident (voice)			
			50.0	0.0	50.0	25.0	H	p	p(overt)
advice	rat → rat	1				0	1.000	1.000	
advice-a	rat-a → rat-a	1				0	1.000	1.000	
wheel	rad → rat	1		1		1	25	1.000	1.000
	rat → rat		1			50	0.000		
	rad → rad	0		1	1	50	0.000	0.000	
	rat → rad		1		1	1	125	0.000	
wheel-a	rad-a → rad-a	1		1		0	1.000	1.000	
	rat-a → rad-a		1			1	75	0.000	
	rad-a → rat-a	0		1		1	25	0.000	0.000
	rat-a → rat-a		1			50	0.000		

10. The fiasco: hand-ranking is easy, but algorithmic-search ranking crashes and burns!

	Hidden	Overt freq	wheel /rad/	wheel /rat/	*Coda Voiced Obs	Ident (voice)			
			0	0	20	0	H	p	p(overt)
advice	rat → rat	1				0	1.000	1	
advice-a	rat-a → rat-a	1				0	1.000	1	
wheel	rad → rat	1		1		1	0	0.5	1
	rat → rat		1			0	0.5		
	rad → rad	0		1	1		20	0	
	rat → rad		1		1	1	20	0	
wheel-a	rad-a → rad-a	1		1		0	0.25	0.5	
	rat-a → rad-a		1			1	0	0.25	
	rad-a → rat-a	0		1		1	0	0.25	0.5
	rat-a → rat-a		1			0	0.25		

- Wrong UR, wrong outputs.
- This is if you take 0 as starting point weights for the Solver.
- If you take a very big starting weight for WHEEL = /rad/, then everything works.
 - This is making it innate that the word for wheel is /rad/, not a hopeful strategy.

11. Why fiasco?

- The summing over hidden structures evidently removes the beautiful **convexity** that makes maxent learning so appealing.
- If you are in the region when WHEEL = /RAD/ is high, then the best ranking of Markedness and Faithfulness is the one that yields final devoicing.
 - IDENT(voice) rightly wants to be high, protecting /rad-a/ and /rat-a/ from undesired random changes.
- If you are in the region when WHEEL = /RAD/ is low, then you are in danger of deriving (from wrong UR) /rat-a/ → *[rata] ‘wheel’
 - Now IDENT(voice) only wants to get out of the way! Being Faithful can only do harm, as it *encourages* the bad outcome.
 - But if IDENT(voice) is near zero, then promoting WHEEL = /rad/ does no good; the UR won’t get enforced.
 - Both say: “Hey, I thought it was your job, so I decided to just nap.”
 - They nap on the couch of a wrong local maximum.
- More generally, we are letting the *violations* of IDENT(voice) be dependent on the values of the UR constraints, a context-dependency that seem responsible for defeating convexity.

12. Efforts to learn hidden structure

- Tesar and Smolensky (2001) *Learnability in Optimality Theory*. An approach called Robust Iterative Parsing; non-stochastic OT.
- Tesar book (2014) *Output-Driven Phonology*, Cambridge University Press.
- Appousidou, cited above
- Gaja Jarosz paper in progress, with a whole new version of OT, evidently best of the lot but not perfect. I would love to try out her system.
 - Jarosz, Gaja. 2015 / in revision. Expectation driven learning of phonology. University of Massachusetts manuscript.

13. The exterminationist approach to hidden structure

- Perhaps hidden structure is more trouble than it's worth?
 - Expansion of hypothesis space
 - Introduction of non-convexities
- Removing the feet has been tried for **metrical stress theory** a number of times:
 - Alan Prince (1983 *LI*, “Relating to the grid”)
 - Gordon, Matthew (2002) A factorial typology of quantity insensitive stress, *Natural Language and Linguistic Theory* 20, 491-552.
- Donca Steriade is an exterminationist w.r.t. **syllables**, a strikingly non-traditionalist point of view, but she has replacement theories in hand for both of their main functions:
 - phonotactics (phonetic cue-based theory)
 - metrical structure (interval theory)
- For **underlying representations**, there is a modest contingent who want to do phonology just with allomorphs, no UR's inferred from allomorphs. Harry Bochner, Luigi Burzio are examples.
- Exterminationists are thinner on the ground in **syntax** (e.g., trees with fewer nodes) but perhaps categorial grammar is an example. Here is an automated-learning-of-syntax paper using this framework:
 - Omri Abend, Tom Kwiatkowski, Nathaniel J. Smith, Sharon Goldwater and Mark Steedman (2017) Bootstrapping language acquisition. *Cognition* 164, pp. 116–143.
- Remember that *complete* extermination of hidden structure is certainly not feasible; there's various stuff I can't imagine we could do without.

PARADIGM UNIFORMITY

14. Goals

- Let us try to integrate the sources of evidence that speakers are predisposed to minimize alternation.
 - don't alternate at all
 - make alternation less phonetically salient

- Inspect the ways that such a bias could be implemented as part of the theory of phonological grammar.

15. Some ancient observations

- Paradigms often shift in the direction of increasing paradigm uniformity.
 - As if Junior refused to believe in the alternations she heard, committed a blunder adopted by her peers.

16. Warm-up exercise: rendaku and [ŋ] in Conservative Tokyo Japanese

- Source: Junko Itô Armin Mester (1997) Correspondence and Compositionality: The Gagyo Variation in Japanese Phonology. In Roca, Iggy, ed. *Derivations and Constraints in Phonology*. Oxford University Press.

Rendaku (sequential voicing in compounds)

tama	‘ball’	teppoo+dama	‘bullet’
tana	‘shelf’	garasu+dana	‘glass shelf’

- Allophone of older speakers of Japanese. They quote the great early-20th-century phonologist Nikolai Trubetskoy, whose name is associated with the word *Grenzsignal*.

Trubetskoi (1949 *Principles of Phonology*, 293): “En japonais il existe entre g et η un rapport de variante combinatoire, g n’apparaissant qu’à l’initiale de mot et η qu’entre voyelles: ici également l’opposition g : η ne peut différencier une paire de mots, mais cette opposition sert à délimiter le mot, g indiquant toujours le début d’un mot.”

“In Japanese there exists between g and η a relationship of combinatorial variant, g appearing only at the beginning of a word and η only between vowels; here as well the opposition g : η cannot differentiate a pair of words, but this opposition serves to delimit the word, g always indicating the beginning of a word.”

Distributional data:

a. Initial g:

✓[g]	*[ŋ]	
geta	*ŋeta	'clogs'
giri	*ŋiri	'duty'
guchi	*ŋuchi	'complaint'
go	*ŋo	('game of) Go'
garasu	*ŋarasu	'glass'

b. Internal ŋ:

*[... g ...]	✓[.... ŋ]	
*kagi	kanji	'key'
*kago	kaŋo	'basket'
*kanŋae	kanŋae	'thought'
*sasageru	sasaŋeru	'give'
*uguisu	uŋuisu	('Japanese) bush warbler'
*tokage	tokaŋe	'lizard'
*igirisu	injirisu	'England'

• Stem-final before suffix; suffix-initial:

oyon	+	oo	'swim-HORTATIVE'	}	stem-final
ton	+	anai	'sharpen-NEG-PRESENT'		
kayoobi	+	ŋa	'Tuesday-NOMINATIVE'	}	suffix/clitic-initial
mikka	+	ŋurai	'approximately three days'		
gorira+no	+	ŋotoshi	'like a gorilla'		

• Alternations in Sino-Japanese compounds, whose members are all *bound roots*:

Bound roots:

PrWd[g...]		PrWd[..... ŋ ...]	
gai + jiN	'foreigner'	kokū + ŋai	'abroad'
go + zeN	'morning'	shoo + ŋo	'noon'
gam + peki	'quay, jetty, wharf'	kai + ŋaN	'sea shore'
gi + kai	'parliament'	shin + ŋi	'deliberation'
guu + zeN	'accidental occurrence'	soo + ŋuu	'meet accidentally'
gen + zai	'currently'	sai + ŋeN	'reappearance'

- How $g \sim \eta$ works in compounds whose members are free stems:

Compounding with g-initial Stem₂: optional VVN

geta	'clogs'	niwa	+ $\{\overset{\eta}{g}\}$ eta	'garden clogs'
goro	'grounder'	pitchaa	+ $\{\overset{\eta}{g}\}$ oro	'a grounder to the pitcher'
gara	'pattern'	shima	+ $\{\overset{\eta}{g}\}$ ara	'striped pattern'
gei	'craft, art'	shirooto+	$\{\overset{\eta}{g}\}$ ei	'amateur's skill'
go	'Go game'	oki	+ $\{\overset{\eta}{g}\}$ o	'Go played with a handicap'

- How $k \sim \eta$ works in compounds with a second-position k-stem:

Compounds involving Rendaku: obligatory VVN

kuni	'country'	yuki	+ $\{\overset{\eta}{g}\}$ uni	'snow country'
kami	'paper'	ori	+ $\{\overset{\eta}{g}\}$ ami	'origami paper'
kaeru	'frog'	gama	+ $\{\overset{\eta}{g}\}$ aueru	'toad frog'
kenka	'fight'	oyako	+ $\{\overset{\eta}{g}\}$ enka	'parent-child fights'
kaki	'writing'	yoko	+ $\{\overset{\eta}{g}\}$ aki	'horizontal writing'
kusuri	'medicine'	nuri	+ $\{\overset{\eta}{g}\}$ usuri	'medical ointment/cream'
kirai	'dislike'	onna	+ $\{\overset{\eta}{g}\}$ irai	'woman-hater, misogynist'

- Exercises:

- formulate an OT analysis of these facts. I suggest *MAP constraints.
- Reconstruct the historical chronology by which this pattern came to be.

17. Spanish verb paradigms as studied by Harris (1973)

- Reference:
- Irregular verbs preserve ancient patterns of Velar Softening in their paradigms.

'do'	'say'	
ha[g]o	di[g]o	1st sing., present indicative
ha[s]emos	di[s]imos	1st plur. present indicative (2nd, 3rd conjugation)
ha[g]amos	di[g]amos	1st plur. present subjunctive

18. Analyzing the alternations

- These can be roughly traced to developments in the history of Spanish.
- N.B. Harris, following contemporary practices, recapitulates these developments as abstract synchronic phonology.

/hak + e + o/	/hak + e + mos/	/hak + e + a + mos /	
hako	hakemos	hakamos	$V \rightarrow \emptyset / _ V$
—	s	—	Velar Softening
g	—	g	Lenition

[hago]	[hasemos]	[hagamos]	$V \rightarrow \emptyset / _ V$
--------	-----------	-----------	----------------------------------

where Velar Softening takes /k, g/ to [s, x]
 Lenition voices intervocalic /k/ to [g].

19. The patterning of regular verbs

“It is an easily observable fact that the same stem-final consonant appears in every form of any regular verb in Spanish, regardless of the desinential vowel that follows this consonant in phonetic representations.”

- So let’s try the opposite order:

/proteg + e + o/	/proteg + e + mos/	/proteg + e + a + mos /	
x	x	x	
∅	∅	∅	Velar Softening
—	—	—	$V \rightarrow \emptyset / _ V$

[protexo]	[protexemos]	[protexamos]	Lenition
			output

- Socrates: characterize the change in Kiparskian terms (feeding/bleeding, etc., opaque/transparent.)

20. Harris’s take

- A problematic case of inconsistent rule ordering, not treatable in contemporary free-ordering theory (< Stephen Anderson).
- Yet he feels guilty about a lurking, very traditional notion: the change in the regulars increased paradigm uniformity.

21. What we might need for OT

- Designation of the base form that rules the roost (cf. research program of Adam Albright).
- Designation of the position, and features, that are regulated.

22. Some further data: paradigm uniformity in first-conjugation verbs

‘mark’	‘pay’	
mar[k]o	pa[g]o	1st sing., present indicative
mar[k]amos	pa[g]amos	1st plur. present indicative (1st conjugation)
mar[k]emos	pa[g]emos	1st plur. present subjunctive

- For Harris, these have an underlying /a/ theme vowel, which drops too late to let Velar Softening apply in the front-vowels subjunctive.

23. What we might need for diachronic explanation

- What caused little Mercedes to want to regularize
- What led her to take particular forms as the basis for extension.

TREATMENT OF PARADIGM UNIFORMITY EFFECTS IN SPE PHONOLOGY

24. The bifurcation

- Inheritance of derived phonological properties: the cycle
- Resistance to acquisition of properties: word-internal boundaries.

25. Cyclic effects

- Already covered by Kie in 201A, but a quick example:
English secondary stresses are (roughly) left-to-right binary, no clash, in the pretonic domain.

Examples from Hayes (1982, *LI*).

àbracadábra	Kàlamazóo		
Lùxipalílla	Hàrdecanúte		
Pèmigewàssett	Àllamakée		
Òkefenókee	Ìllilouétte		
Nèbuchadnèzzar	Màttamusquéet		
pàraphernália	Àntigonísh		
Kìlimanjáro	Gàllipolís		
Pòpocàtepétl	Òkalòacóochee	Àpalàchicóla	Àntanànarívo
Hànamànióa	ìpecàcuána	ònomàtopóeia	hàmamèlidánthemum

This is not respected in suffixed forms, where the principle seems to be inheritance, modulated by the need to avoid clashes and initial lapses:

morphologically derived long words typically do not display the pattern of secondary stress found in monomorphemic words: compare *sublimináli* with *Òkefenókee*, *demòcratizáti* with *Àpalàchicóla*, and *Macàssaré* with *Gàllipolís*.

- Analytic possibility: do not foot what is already footed on a prior cycle, but do resolve certain clashes (*democratization*) and the resulting lapses (*specificity*).

26. Boundary effects: the distribution of preantepenultimate stress

- There are no stems whatever ending in stressed plus three stressless: “Hi, I’m *['pæmələnə]”
- Within the learned vocabulary, a few affixes give rise to the pattern.
 - Socrates: while we’re at it, look for possible Paradigm Uniformity effects in these data.

abominable

applicable

communicable

estimable

inalienable

incalculable

inextricable

innumerable

inseparable

interminable

inviolable

irremediable

navigable

permeable

tolerable

venerable

actionable

enviable

fashionable

fissionable

impressionable

knowledgeable

objectionable

perishable

practicable

questionable

reasonable

seasonable

serviceable

variable

amiable
amicable
formidable
indefatigable
malleable

caricature
temperature
literature

communicative
palliative
speculative
cumulative

accuracy
adequacy
advocacy
candidacy
celibacy
confederacy
degeneracy
delicacy
immediacy
intimacy
intricacy
legitimacy
literacy
obstinacy

occupancy
militancy
hesitancy
relevancy
irrelevancy
residency
presidency
expediency
incompetency
constituency

idiocy

- With productive suffixes, pre-antepenultimate stress seems rather normal and possible in new words:

<i>-ing</i>	<i>monitoring, jettisoning</i>
<i>-eth</i>	<i>seventieth</i>
<i>-ish</i>	<i>Madison-ish</i>

- *SPE*: the productive suffixes are treated with “#”. “Readjustment rules” apply.
 - Rule 1: [] → [# #]
 - Rule set 2: X #] ation → X] ation; etc., for the less-productive affixes.
 - Stress rules apply in domains bounded by # #.
 - In translated form (prosodic structure), this is still a living analytic option, see e.g. Peperkamp, S. (1997). *Prosodic Words*. HIL dissertations 34. The Hague: Holland Academic Graphics.

27. Returning to the bifurcation

- These receive a uniform treatment under Paradigm Uniformity constraints.

28. Something that emerges from this discussion

- Paradigm Uniformity is sensitive to the paradigm involved; i.e. we may need to be quite specific about the morphological relations present.

THE PRINCIPLE OF PHONOTACTIC LIBERALITY

29. Trying to express the principle

- If you inspect the inventory of monomorphemic forms, you will get a rather strict phonotactics.
- But when forms occur in paradigms, a wider variety of legal forms emerges.

30. Sources of richer phonotactics in paradigms

- Suppressed phonology
 - like *'monitoring*, not **mo'nitoring*
 - Socrates: find another example on this handout.
- Overapplied phonology, like *sublimi'nality*
- Mere concatenation, like
 - “Hello, my name is Bill *[trebd]. I and all the other *[trebdz] are very pleased to meet you.”
 - Yet: *rubbed, dubbed, ribbed*, etc.
 - Socrates: what constraints could account for this pattern, including Paradigm Uniformity?

31. A classic example from the urtext of Paradigm Uniformity in OT

- Laura Benua (1997) *Transderivational Identity: Phonological Relations between words*. U. Mass. dissertation.
- Epenthesis:
 - Generally, words in Tiberian Hebrew do not end in consonant clusters.
 - There are a tiny number of lexical exceptions,
 - plus a larger class of systematic exceptions.
- Example:

Epenthesis in Tiberian Hebrew is demonstrated in (85) with the monomorphemic word [sēpher] 'book', which is related to the input root /sipr/ (compare [siɸrī] 'my book', in which the root's consonant cluster surfaces intact in a heterosyllabic parse).
- Jussives are formed by final vowel loss from imperfective base, yet often there is no epenthesis:

Jussive Truncation

	<u>Imperfective</u>	<u>Jussive</u>	
a.	yiš.bē	yišb	'take captive'
	yiɸ.tē	yiɸt	'be simple'
	yēš.te	yēšt	'drink'
	yēβ.ke	yēβk	'weep'
	yiš.te	yēšt	'drink'
	yaš.qe	yašq	'cause to drink'
b.	yiγ.lē	yi.yel	'uncover'
	yiβ.ne	yi.βen	'build'
	tiɸ.nē	tē.ɸen	'turn'
	yiβ.zē	yi.βez	'despise'
	yiš.ʃē	yi.ʃaʃ	'gaze'
	not attested	yi.ħad	'rejoice'

- Figure out an analysis in Classical OT.