Linguistics 201A | Winter 2017
Phonological Theory I | B. Hayes

Class 18, 5/27/2020: Contrast in Phonology II

1. Assignments

- Read:
  - Stanton, Juliet (2018) Environmental shielding is contrast preservation. *Phonology* 35:39–78.
  - On course web site.
- Feel free to discuss progress/problems with your term project.
  - Office hours W, F at 2 and by appt.
- Handout-session with me before the middle of Finals Week, Wed. 6/10.

2. Summary of last time

- Flemming, following phoneticians, laid out (and formalized) the triple conflict of dispersion:
  - lots of categories is good, to make language possible
  - lots of categories is bad because they must be perceptually close to one another
  - extreme categories are articulatorily expensive
- He laid out a radical proposal that optimizes not over forms but over systems, with three constraint types.
  - *HIGH EFFORT (number)
  - MINDIST=Dimension:Number
  - MAXIMIZE CONTRASTS (number)
- An alternative point of view is diachronic/acquisitional: these forces create dispersed phonologies, but the phonological grammar is "unaware" of these patterns — deadpan, literal. (Boersma, Blevins 2004 book, Elliott Moreton, others).

MORE ON THE DIACHRONIC/ACQUISITIONAL PERSPECTIVE:
DISPERSION THROUGH CORRECTED PERCEPTION ERROR

3. The Boersma/Hamann strategy

- Reference:
  - Boersma, Paul and Silke Hamann (2008) The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25: 217-270
- The child is assumed to make "correctible" perception errors.
  - I.e. she knows what word she is hearing, but her perceptual apparatus returned a different phoneme.
  - Example: hearing *say* [se] as [si].

- These errors lead to small shifts in the phonetic targets, which make it harder to make this error in the future.

4. Goal here

- Take this out of the armchair (as Boersma and Hamans do) with a model.
- But using Flemmingian Parabolic Generative Phonetics, rather than the not-spreadsheetable Boersmian Exploded Constraint Family Generative Phonetics.

5. A Scenario

- The target language has just two high vowels, [i] and [u].
- They differ on a scale that ought to be F2, but for convenience here will be a fictional scale ranging from 0 to 200.
- A fictional adult:
  - He says the utterances [li] and [lu], alternatingly, in the presence of a baby named Junior, every 30 seconds.
  - He is in fact yelling, individually, at Junior's obstreperous older siblings *Lee* and *Lou*, and Junior, in her infant-seat, is a witness.
  - So any speech perception errors that Junior makes are *correctable* errors.
- Like all children, Junior follows this acquisition strategy:
  - If you erroneously hear [li] for intended [lu], raise the target for [i].
  - If you erroneously hear [lu] for intended [li], lower the target for [u].
  - In other words, shift targets to make errors less likely to be repeated.

6. Refinements and simplifications

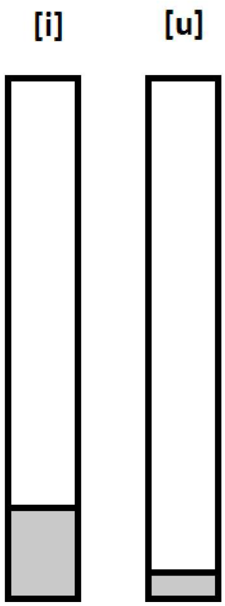
- The adult is assumed to produce [i] and [u] each with a Gaussian (Flemmingian) probability distribution.
- Having run out of time, I'm using the child's perceptual distributions to approximate the adult production distributions, which probably is introducing distortions but not fatal ones.
- We will say that Junior makes a "wrong guess" if she assigns a probability greater than .5 to the wrong vowel.

7. How might the child do her guessing? Forward probabilities

- She has a Gaussian that represents a theory of the parental distribution of F2 values for both [i] and [u].
- Following the Boersmian idealization, we use not continuous probability but slices; integers along the arbitrary scale of 1-200.
- It is then a matter of simple look up: P([i], 130), P([u], 130)
- These are **forward probabilities**: the child's theory of adults production grammar.

8. Going from forward probabilities to perception probabilities

- We need some reasoning, probably Bayesian, which I'm not sure of (you can help)
- Here is the relevant box diagram for the value 130.



  - ➢ Gray regions: fractions of all [i], or all [u], that land in the 130 time slice in assumed adult distributions.
  - ➢ Bars are equal width, i.e. equal number of total [i] and [u]
  - ➢ Then: probability that 130 is [u] = ratio (right gray box / total of two gray boxes)
  - ➢ This is standard reasoning; infer backward probabilities from forward probabilities.

9. Child's assumed response to the data

- If, with her grammar, she makes a known perception error, then shift the target of the wrong guess up ([i]) or down ([u]) to make this error less likely.
- We will say a wrong guess is defined by assigning a probability greater than .5 to the wrong candidate.
- We can pick a modest value to change the target.

10. Go to spreadsheet that implements this

11. Lots of open questions

- What if we keep a constant distribution of adult outputs?
  - ➢ This is what I would have done, with more time.
  - ➢ Does the child's system frequency-match (same Gaussians and targets) or overshoot (biased learning)?
- Adjusting for differing vowel frequencies

- These are the box widths in (8) above.
- This is probably trivial.

- Continua with multiple targets
  - If you move target for [e] downward to avoid confusion with [i], you increase likelihood of confusion with [æ] — which will then move downward itself, once the model does its work.

- Abandoning the .5 cutoff
  - Would something gradient (shift depends on size of error) work more effectively?
  - Boersma mentions this.

- Folding in articulatory difficulty
  - Following our earlier assumptions, the child eventually learns to govern her own productions to match (modulo vowel normalization) those of her parents
  - In doing this, she would confront the barrier of articulatory difficulty as she attempts extreme targets for [i] and [u].
  - Perhaps the child characteristically fails to reach her own targets???

- The Hyperspace Effect
  - Reference: Johnson, Keith, Edward Flemming, and Richard Wright. "The hyperspace effect: Phonetic targets are hyperarticulated." *Language* (1993): 505-528.
  - UCLA undergraduates asked to click on an array of vowels varying smoothly in F1 and F2 to match the vowels of their language pick *extreme* targets — perhaps reflecting the inner ideal they learned while cautiously learning to disperse as children?

- Drag chains
  - This is standard historical lore: sounds move, and other sounds fill the gap.
  - E.g. historical Swedish, *u → y, o → u
  - So something must make sounds feel free to drift in directions not block by perceptual difficulty.

- Who, other than Boersma, has done this for real?
  - I think Boersma's ideas are great, but I also feel this is an ideal zone for the application of existing math (mainstream probability theory and Bayesianism)

12. Summing up the acquisitional/historical account of dispersion: four phenomena

- The "**excluded center**" in phonotactics
  - This we have just seen: acquisition creates dispersion, at the cost of effort.

- **Enhancement** of outputs in alternation.
  - Probably the same thing.

- The "**neutralization of despair**" in poor-contrast conditions.
  - Here, we appeal to natural sound change.
  - Sound change must be gotten away with — it is weird to sound different from everyone else.

  - ➢ If there is FAITH$_{\text{me-them}}$, it plausibly is governed by the P-map, and thus neutralizing sound change will be in non-salient conditions.

- The **fortition of rescue**, if it is a valid principle
  - ➢ This is perhaps a property of the clearer speech registers?
  - ➢ It ties into phonetic research on clear speech, notably the degree of teleology/wisdom it embodies.

JULIET STANTON'S WORK ON NC … NC DISSIMILATION

13. Source

- Stanton, Juliet (2019) Constraints on contrast motivate nasal cluster dissimilation. *Phonology* 36: 655–694.

14. The phenomenon

- All over the world the sequence NC … NC tends to be avoided.
- Two examples:

Ngaju Dayak: deletion of the NC_1 *oral component* (Blust 2012: 371–372)

i.	/maN-bando/	→	[mamando]	'turn against'
cf.	/maN-bagi/	→	[mambagi]	'divide'
ii.	/maN-gundul/	→	[maŋundul]	'shave'
cf.	/maN-gila/	→	[maŋgila]	'drive crazy'

Gurindji: deletion of the NC_2 *nasal component* (McConvell 1988: 138)

i.	/kaɲɟu-mpal/	→	[kaɲɟupal]	'across below'
cf.	/kajiɾa-mpal/	→	[kajiɾampal]	'across the north'
ii.	/kanka-mpa/	→	[kankapa]	'upstream'
cf.	/kani-mpa/	→	[kanimpa]	'downstream'

- This is very odd! Typically such effects reflect a single feature.
  - ➢ Stanton cites Will Bennett's recent book, a survey of long-distance dissimilation.

description	features involved	*n*	example language
C-place	[+labial], [+coronal], [+dorsal]	42	Akkadian
nasal	[+nasal]	2	Takelma
laryngeal features	[+constricted glottis], etc.	29	Aymara
continuancy	[+continuant]	5	Chaha
liquids/rhotics	[±lateral]	22	Latin
sibilants	[+strident]	4	Nkore-Kiga
voicing	[−voice]	29	Kinyarwanda
NC sequences	[+nasal] [−nasal]	21	Gurindji

- Stanton tends to be very noncommittal about whether NC is treated as a cluster or a single prenasalized segment; it seems not to matter for typology.

15. Stanton's basic hypothesized mechanism for why NC … NC is hard

- Coarticulatory nasality
  /NCVNC/ → [NCṽNC]
- Tight velum-slalom at the initial C makes it realized in an inferior way.
- and so /NCṽNC/ and /NṽNC/ get too close for comfort.
- The typically observed response is the Neutralization of Despair.
- The neutralization favors the articulatorily easier [NṽNC].

16. Stanton covers a lot of detail I: right-side NC is a better trigger than right-side N

- Survey:
  - 43 languages of survey repair NCVNC, but not NCVNV.
  - 21 languages repair only NCVNC.
  - The 3 cases of repairing only NCVNV can be explained away.
- The mechanism:
  - Coda nasals (perhaps also: postvocalic prenasalization-bits) have lower velum position (a standard of the phonetic literature; cf. initial strengthening of /n/ to [d] in Korean)
  - And so the vowel has more coarticulatory nasality.
  - And so the C of the first NC has a worse slalom-turn and is in greater perceptual peril.

17. Stanton covers a lot of detail II: triggering in NCVNC if and only if triggering in NCVN#

- Again, apparent exceptions can be explained away.
- This is classical phonology: a coda is a coda, whether it arises preconsonantally or word-finally.

18. Stanton covers a lot of detail III: triggering is sometimes stronger in NC_1VNC_2 if C_2 is voiceless

- This too, reflects phonetics, with typically greater nasality in this environment.
- In this case, Stanton downloaded a lot of Yindjibarndi phonetic data from UCLA's archive and measure.

19. Stanton covers a lot of detail IV: triggering is sometimes stronger in NC_1VNC_2 if C_1 is voiceless

- This is pretty straightforward, reflecting the more salient release information available in voiceless consonants.

20. Stanton covers a lot of detail V: triggering is sometimes stronger in NC_1VNC_2 if C_1 has a "quiet" place of articulation

- Notably, bilabials have quiet releases because there is no chamber within which the burst noise can resonate.
- Ngaju Dayak has place asymmetries, which are stochastic.
- They track release noise, measured by Stanton from speakers reading the Bible in Ngaju Dayak on line.

FURTHER DEVELOPMENTS IN DISPERSION THEORY: *MERGE

21. Ur-source

- Work of Jaye Padgett, notably
  - Padgett, Jaye (2003). Contrast and post-velar fronting in Russian. *NLLT* 21. 39–87.
- Here, working off the use made of it by Stanton in her NCVNC paper.

22. MERGE per Stanton (p. 660)

*MERGE

No output word has multiple correspondents in the input.

23. In other words

Neutralization is penalized.

24. Purpose

- It replaces the original MAXIMIZE CONTRASTS constraint used in the ur-research of Flemming.

25. The full-dress version of tableaux in grammars using Merge, from Padgett (2003)

- Padgett, Jaye. "Contrast and post-velar fronting in Russian." *Natural Language & Linguistic Theory* 21, no. 1 (2003): 39-87.

pi_1 $pɨ_2$ pu_3	*Merge	Ident(Color)
a. pi_1 $pɨ_2$ pu_3		
b. $pi_{1,2}$ pu_3	*	*

- The indices are not the correspondence indices of McCarthy and Prince, but rather indicate what is derived from what.

26. Translating a Flemming tableau

	*HIGH EFFORT	MINDIST=F1:2	MAXIMIZE CONTRASTS
[ɨ]			✓
[i]	*		✓
[u]	*		✓
[i-ɨ]	*	*	✓✓
[ɨ-u]	*	*	✓✓
[i-u]	**	*	✓✓
[i - ɨ - u]	**	**	✓✓✓

- With MERGE, we add to the input row the impossible candidates that have, by hypothesis, been realized identically to the correct candidate.

	*HIGH EFFORT	MINDIST=F1:2	MERGE
[ɨ] + [i],[u]			**
[i] + [ɨ],[u]	*		**
[u] + [i], [ɨ]	*		**
[i] in a system with [ɨ] + [u]	*	*	*
[ɨ] in a system with [u] + [i]	*	*	*
[i] in a system with [u] + [ɨ]	**		*
any vowel in a system with [i] + [ɨ] + [u]	**	**	

- I hope I have filled in the candidates and MERGE violations correctly.
- It can be seen that the MERGE violations map as an exact inverse of the MAXIMIZE CONTRASTS violations, so the factorial typology will be exactly the same.

27. The theory applied to Stanton's first empirical case

- Stanton's key sequence of reasoning:

- ➢ The phonetics determines the degree of nasality of the V in NCVNC.
- ➢ That nasality determines the slalom-difficulty of the first C.
- ➢ That difficulty determines the degree to which the first NC is perceptually similar to N.
- ➢ That determines the ranking/weighting of a MINDIST constraint.

28. Settting up the MINDIST constraints I: defining an orality scale for vowels

- Set up idealized phonetic degrees of orality in vowels, and call them V_{Oral1}, V_{Oral2}, V_{Oral3} etc., thus:

$Oral_x$ *value*	*referent*	*example*
1	V / __N_{coda}	amanda
2	V / __N_{onset}	amana
3	other	amada, amada

29. Settting up the MINDIST constraints II: deploying the orality scale

- Let these categories serve as right-side environments of MINDIST constraints, stated thus.

(4) a. MINDISTN–NC(V_{Oral2})
Assign one violation for each contrast between N and NC in which the following vowel does not belong to category 2 or higher on the Oral scale.

b. MINDISTN–NC(V_{Oral3})
Assign one violation for each contrast between N and NC in which the following vowel does not belong to category 3 or higher on the Oral scale.

30. Now the typology emerges smoothly

- It all depends on which MINDIST constraints outrank MERGE.
- English is very liberal (Stanton gives *commend* / *compend*; 200 NCVNC hits on my little program)

(5) *English:* *MERGE *undominated, no nasal cluster dissimilation*

$/amanta/_i$ $/ambanta/_j$	*MERGE	MINDIST N–NC(V_{Oral2})	MINDIST N–NC(V_{Oral3})
☞ a. $[ama_1nta]_i$ $[amba_1nta]_j$		*	*
b. $[ama_1nta]_{i,j}$	*!		

- Ngaju Dayak is intermediate: only NCVNC banned, not NCVNV

(6) *Ngaju Dayak:*
nasal cluster dissimilation only when the second N is a coda

		MINDIST N–NC(V_{Oral2})	*MERGE	MINDIST N–NC(V_{Oral3})
a.	/mambando/$_i$ /mamando/$_j$			
	i. [mama$_1$ndo]$_i$ [mamba$_1$ndo]$_j$	*!		*
	☞ ii. [mama$_1$ndo]$_{i,j}$		*	
b.	/maneŋen/$_i$ /mandeŋen/$_j$			
	☞ i. [mane$_2$ŋen]$_i$ [mande$_2$ŋen]$_j$			*
	ii. [mane$_2$ŋen]$_{i,j}$		*!	

- Luganda is strictest:

(7) *Luganda:*
nasal cluster dissimilation when the second N is a coda or an onset

		MINDIST N–NC(V_{Oral2})	MINDIST N–NC(V_{Oral3})	*MERGE
a.	/mumba/$_i$ /mbumba/$_j$			
	i. [mu$_1$mba]$_i$ [mbu$_1$mba]$_j$	*!	*!	*
	☞ ii. [mu$_1$mba]$_{i,j}$			*
b.	/nimi/$_i$ /ndimi/$_j$			
	☞ i. [ni$_2$mi]$_i$ [ndi$_2$mi]$_j$		*!	*
	ii. [ni$_2$mi]$_{i,j}$			*

31. The general ranking principle is straightforward

- Ban contrasts where MINDIST outranks MERGE.
- Allow contrasts where MERGE outranks MINDIST.

SOME COMMENTS ON THIS WORK/FRAMEWORK

32. Caution

- You are getting non-expert opinion; I've never worked in this framework and am underread.

33. I am happy

- Whenever I see work that extracts broad, meaningful patterns from seemingly petty little details, I am intellectually happy.
- Kiparsky (*LI* 1977), quoting Jespersen (1933): "life consists of little things, the important matter is to see them largely"

34. I am baffled/frustrated

- How does this really work? The tableaux are illustrative/schematic, leading to Bruce-bafflement.
- Specifically:
  - A *huge number* of neutralizations is often possible with any derivation, particularly in orthodox OT with a Rich Base.
  - To define the phonotactics, OT lets in all the possible-but-crazy things, and kills them all off with a deft constraint-ranking.
  - What should I be putting in the left column of the tableau?

35. Example with *cat*

- Feasible URs:
  - /kæt/
  - /kætq/
  - /kætqq/
  - /kætqqq/
  - /kætqqq/
  - /kætqqqq/
  - (infinity)
- All but the first are duly killed off by *[q] > MAX.

36. Perhaps this is harmless?

- What we are interested in is whether the winner has a **subset or superset** of the MERGE violations of the loser — the infinities of irrelevant comparisons cancel out.
- If so, I'm still uneasy in the absence of some solid reassurance (mathematical? computational?) that this method works *at scale*.

37. Where we might turn for guidance

- Stanton's paper refers us to Padgett (2003), cited above.
- Stanton's dissertation refers us to unpublished work by Edward Flemming
  - Flemming, Edward. 2008b. The Realized Input. Ms., Massachusetts Institute of Technology, Cambridge,
- A quick look at these does not really provide answers to my puzzlement.

38. One more comment about the Stanton paper

- Everything that she discovered (typology, phonetics, phonetic chains of reasoning) might be grist for the mill of diachronist/acquisitionist dispersion approaches.