Class 19 (Week 10, T)
Induction II: constraints & features

To do
- ☐ Work on your project.
- ☐ Because we're having class right now, during my usual office hours, we should negotiate some office hours for later in the week.
- ☐ While we're at it, let's negotiate some for next week too.

Overview: What if we aren't born with a constraint inventory, or even a feature set? We'll continue our tour of some proposals for constraints, then talk about features.

1. Flack (2007): inducing a constraint from perceptual experience

- There are languages that prohibit [p] specifically in word-initial position: *#P
  - Initial [p] has particularly short VOT, and it's more variable than initial [b]'s
  - Difference in maximum burst intensity for initial [p] and [b] is smaller than for other voiceless-voiced pairs (p. 122)

- To produce an instance of a category ([p], [b], [t], etc.) in a context, speaker samples values for various phonetic dimensions from stored distributions centered on prototype
- In perception, listener must guess the category
  - Some noise is added: perception is imperfect
  - Rather than Bayes' rule as in Kirby (2013), finds the closest prototype
- Listener gets feedback on accuracy
  - Allows listener to update prototypes
  - Listener also **stores accuracy rate** for each category, perhaps over a moving window of the past *n* tokens (here, *n*=400)
    - Specifically, **hit rate** and **false alarm rate**
    - Does anyone know these terms?

- Hit rates for each consonant as learning proceeds over time:

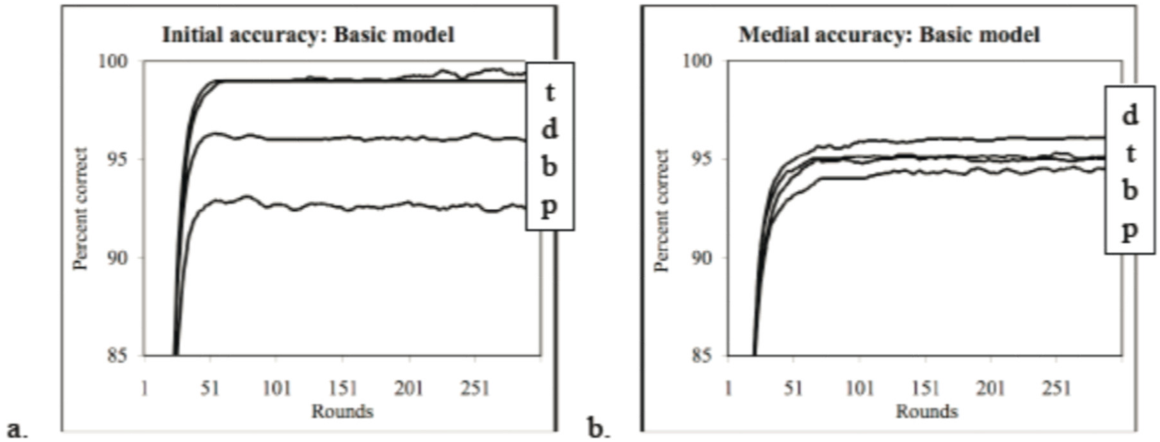

Figure 7. Model accuracy for each initial (a) and medial (b) consonant, averaged across 20,000 simulations of 300 rounds each. The lines represent moving averages across 15-round windows.

(p. 141)

- Important point about phonologization
  - ▪ Once *#P has been promoted high enough, the learner gets no experience of initial [p]!
  - o But they do still have one important piece of information listed above—can you guess what?
- The learner's rule for inducing a constraint:

(117) Some segment *x* is perceptually difficult in some context *Context*$_Z$ if either:

a. Accuracy(*x*/*Context*$_Z$) < *threshold*

and

Accuracy(*x*/*Context*$_Z$) < Accuracy(*y*/*Context*$_Z$) ➔ Constraint *x/*Context*$_Z$

This difference must be significant ($\alpha = 0.01$).

b. Accuracy(*x*/*Context*$_Z$) < FalseAlarm(*x*/*Context*$_Z$) ➔ Constraint *x/*Context*$_Z$

(p. 160)

- ▪ Where "accuracy" means hit rate
- ▪ If there is no hit rate, because the sound never actually occurs in that context, treat it as 0.
- o So how would this work for a language with no initial [p]? Let's draw a possible confusion matrix.

- Results: both a learner of simplified French (has initial [p], but it is perceptually difficult) and a learner of simplified Cajonos Zapotec (no initial [p]) learn *#P in nearly all runs

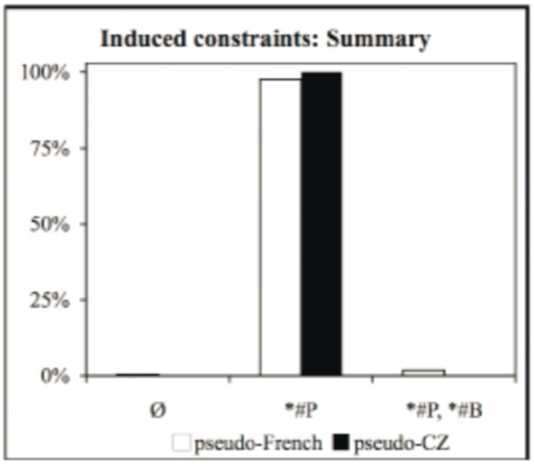

Figure 23. **Constraints induced in each of 250 simulations of 40,000 rounds each.**

(p. 173)

2. A selection of other approaches that we won't have time to cover in depth

- Hayes (1999): as we saw previously (when talking about phonologization), generate lots of constraints according to a set of templates, and then select the ones that match the articulatory-difficulty map well (high accuracy in saying which of two cells in the map is harder), with a bias favoring simpler constraints

- Boersma & Pater (2007): in Harmonic Grammar, construct ***positive*** constraints for every property that the observed form has (as well as some other constraints, including negative ones)
  - e.g., on observing Canadian English [ʔʌɪs] 'ice', construct these, among many others:

(17)	*Observed structure*	*Constructed constraint*
a.	D [-low]	RAISED DIPHTHONG: A diphthong must be [-low] (Assign a reward of 1 to each diphthong that is [-low])
b.	D C [-low][-vce]	RAISED/VOICELESS: A diphthong preceding a voiceless consonant must be raised (Assign a reward of 1 to each diphthong that is [-low] that precedes a voiceless consonant)

(p. 4)

  - Discuss: We've mentioned earlier some problems that having positive constraints could cause. How do the above constraints get around them?

  - Pater (2014) proposes something similar for the same case, but now without positive constraints

- Moreton (2010): explore infinite space of possible constraints with evolutionary algorithm
  - Every subpart of every possible representation is a constraint
  - Start with a random set of constraints
  - Error-driven: if current grammar selects a candidate that doesn't match the observed true winner…
    - constraints that favor observed forms (correct winners) are allowed to breed.
    - breeding = combine two constraints to produce a new, offspring constraint with aspects of each parent. Offspring can also mutate.

- Pizzo (2013): Inducing constraints to handle alternations (Turkish vowel harmony and devoicing)
  - On making an error, create a constraint at random, according to certain templates
  - Can be faithfulness or markedness
  - Must penalize some property on which the spurious winner differs from the observed winner
  - The researcher can set parameters for how much stem-faithfulness and tier-markedness constraints show be allowed/favored

- Alderete, Tupper & Frisch (2013): Connectionist model of OCP-Place in Arabic roots

3. Inducing features

o Discuss: What are features for, anyway? What would we take as evidence for or against the claim that features are universal (or even innate)?

4. Mielke (2004): a cross-linguistic survey

- A lot of rules defy feature analysis.
  - E.g., ChiMwiini palatalization before suffix *-iiɫ-*:

(2) p t̪ t → s
k → ʃ
b d̪ d g → z / [+nasal]__
ɫ → z

(p.3)

- 561 languages, ~17,000 rules or phonotactics, 6077 total (distinct) classes referred to by rules in those languages
- Three feature theories—how well do they capture these 6077 classes?

Feature System	Characterizable (Natural)		Noncharacterizable (Unnatural)	
Preliminaries	3640	59.90%	2437	40.10%
SPE	4313	70.97%	1764	29.03%
Unified Feature Theory	3872	63.72%	2205	36.28%
ANY SYSTEM	4579	75.35%	1498	24.65%

Table 5.3. The ability of three feature systems to characterize 6077 phonologically active classes with a conjunction of distinctive features

(p. 190)

- What if we allow some set operations other than just intersection?

Best analysis	*Preliminaries*		*SPE*		Unified Feature Theory	
Natural (feature conjunction)	3640	59.9%	4313	71.0%	3872	63.7%
Disjunction (2 classes)	1443	23.8%	1248	20.5%	1266	20.8%
Subtraction (2 classes)	59	0.97%	71	1.17%	94	1.55%
Disjunction (3 classes)	233	3.83%	201	3.31%	205	3.37%
Disjunction (4 classes)	64	1.05%	56	0.92%	67	1.10%
Disjunction (5 classes)	17	0.28%	21	0.35%	17	0.28%
Disjunction (6 classes)	0	0.00%	4	0.07%	5	0.08%
Disjunction (7 classes)	1	0.02%	0	0.00%	0	0.00%
Disjunction (8 classes)	0	0.00%	0	0.00%	1	0.02%
Disjunction (9 classes)	0	0.00%	1	0.02%	0	0.00%
Unnatural (even w/disjunction)	620	10.2%	162	2.67%	550	9.05%

Table 6.2. The ability of three feature systems to characterize 6077 phonologically active classes with a conjunction, subtraction, or disjunction of distinctive features

(p. 193)

  - Example of "unnatural even with disjunction": only central vowels, but not front or back

- Are the theories at least doing better than a null hypothesis?
  - Generate 6077 random sets of phonemes (with same distribution of sizes) and see how many are captured by each theory

Best analysis	*Preliminaries*		*SPE*		Unified Feature Theory	
Natural (feature conjunction)	342	5.63%	467	7.68%	270	4.44%
Disjunction (2 classes)	1718	28.3%	1994	32.8%	1745	28.7%
Subtraction (2 classes)	9	0.15%	17	0.28%	11	0.18%
Disjunction (3 classes)	948	15.6%	1160	19.1%	939	15.5%
Disjunction (4 classes)	624	10.3%	774	12.7%	630	10.4%
Disjunction (5 classes)	349	5.74%	456	7.50%	352	5.79%
Disjunction (6 classes)	247	4.06%	292	4.81%	246	4.05%
Disjunction (7 classes)	107	1.76%	126	2.07%	121	1.99%
Disjunction (8 classes)	29	0.48%	29	0.48%	48	0.79%
Disjunction (9 classes)	8	0.13%	3	0.05%	16	0.26%
Disjunction (10+) or error	241	3.97%	290	4.77%	400	6.58%
Unnatural (even w/disjunction)	1455	23.9%	469	7.72%	1299	21.4%

Table 6.4. The ability of three feature systems to characterize 6077 randomly-generated classes with a conjunction, subtraction, or disjunction of distinctive features

(p. 195)

- Mielke's take: We can't just write off these "unnatural classes" as "oddities", because…
  - A lot of them recur in multiple languages, just like "natural" classes are supposed to
  - The frequency distribution of "natural" vs. "unnatural" is not bimodal or well separated
  - Taking the distribution for SPE, which had the best results overall (also shows the best separation):

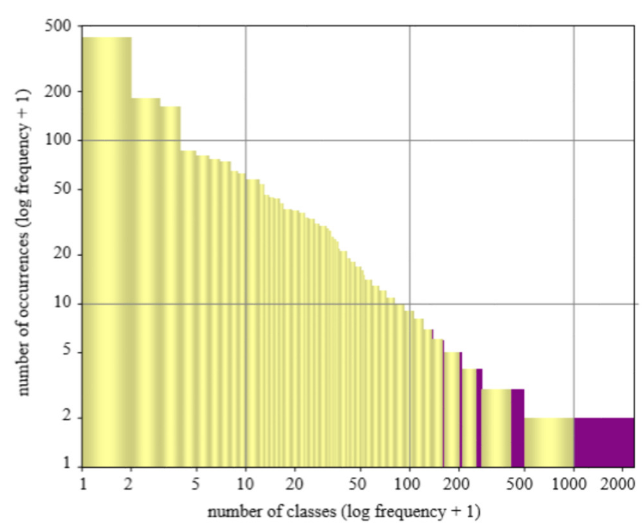

Figure 6.6. The distribution of frequent and infrequent natural and unnatural classes (SPE)

(p. 204)

- So why do some classes recur, if it's not because of an innate feature inventory?
  - Some phonetic effects just naturally involve classes of sounds, like vowels near nasal consonants tend to be a little nasalized, which can then get phonologized.
- Shared phonetic properties do seem to matter to the learner, though:
  - Schaffhausen Swiss German: originally, o → ɔ / __r
  - Seems reasonable as a phonologization of a phonetic effect of [r] on [o]
  - The rule has gotten generalized differently in different parts of the area:

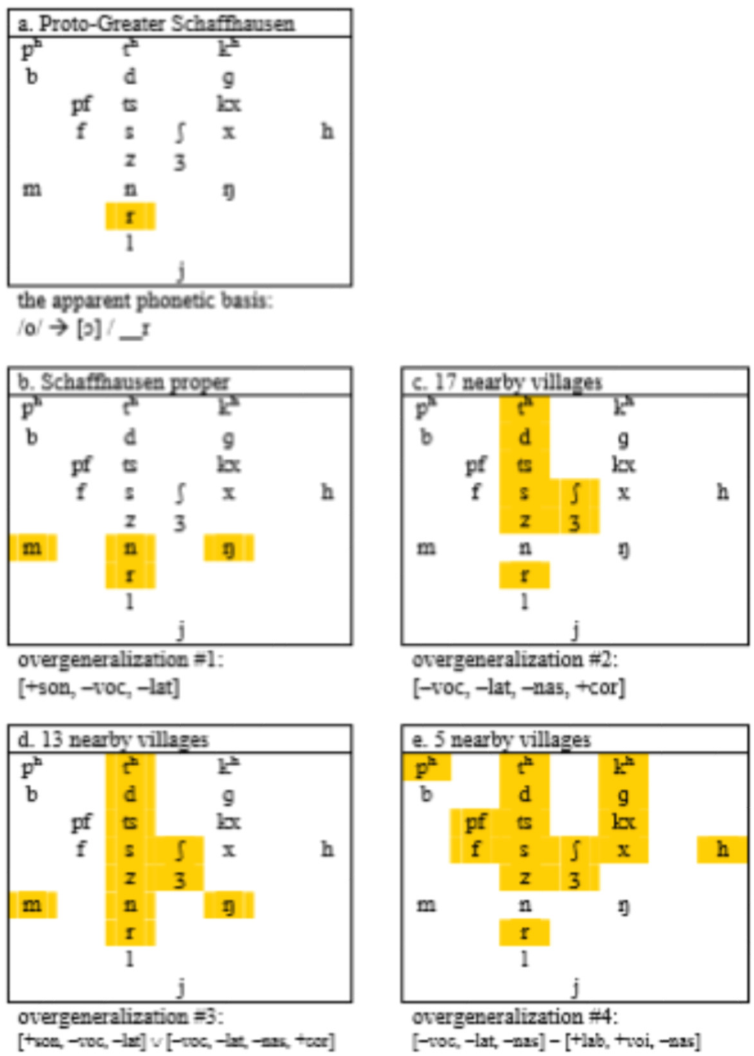

Figure 4.7. Generalization of the conditioning environment for a sound pattern in Schaffhausen Swiss German

(p. 108)

- Mielke concludes that learners construct features in response to learning data
  - We can give them names, but a feature F is just defined by what set of sounds are +F and what set are –F
  - What remains to be proposed is an explicit algorithm for detecting groups of sounds that pattern together, and inducing features from it

5. Flemming (2005): putting features into the grammar

- Discuss: In OT, there is no phoneme inventory. What work was the phoneme inventory supposed to do in rule theories, and how does an OT grammar accomplish that work?

- In a similar move, Flemming proposes getting rid of the feature set, and shifting its responsibilities to the constraint inventory.

- An issue Flemming raises for natural classes: Suppose you have a vowel inventory /i,e,a,o,u/ and you want a rule-based grammar that deletes /i,a,u/ __ V. What could you do? (no curly brackets allowed)

- Then if there are no such rule-based languages, is there a way to use feature theory to rule them out?

- How would we analyze the language in OT?

- Flemming's proposal: if we want to rule out this language, it has to by disallowing the constraints needed to capture it.
  - It won't suffice to just say that constraints can only refer to natural classes (why?)
  - For example, "[i]f labials and coronals never pattern together as a natural class [*e.g., in post-nasal voicing*], it must be because there are no constraints that render them [*but not, say, velars*] marked in the same context." (p. 12 of ms. version)
- If, on the other hand, there is a good reason for a bunch of constraints to exist, like *NAS-APPROX, *NAS-FRIC, *GEMINATE_NASAL (Lithuanian *n*-deletion), then it will seem as though {approximants, fricatives, nasals} is acting as a class
  - Flemming goes through typological data to justify the three constraints (plus *NAS-[h]
    - i.e., there are languages with one of the constraints high-ranked, but not the other two

  - General principle: “sounds can pattern together as a natural class if they violate markedness constraints in the same environment, so given constraints *XA and *XB, A and B can form a natural class” (p. ?)
- “Classhood” is contingent
  - {approximants, fricatives, nasals} can pattern together after nasals specifically, because of the constraint set
  - But we don’t expect them to pattern together in any other environment necessarily
  - I think this is a difference in predictions from Mielke—we can discuss if time.
- How to get subtraction: Markedness1 >> Markedness2
  - Pharyngealization ([RetractedTongueRoot]) spread in Palestinian Arabic
  - Spreads rightward until it hits a high front vowel, a front glide, or a palato-alveolar C
    - all of those are ([+high, -back])

(33)	a.	tˤuubɑk	‘your blocks’		tˤwɑɑl	‘long (pl.)’
		bɑllɑɑsˤ	‘thief’		ʔɑbsɑtˤ	‘simpler’
	b.	tˤiinak	‘your mud’		sˤɑjjad	‘hunter’
		ʕɑtˤʃaan	‘thirsty’		ðˤɑddʒaat	‘type of noise (pl.)’

(p. 34)

  - What’s the class of sounds that pharyngealization spreads to?

  - How could we capture that in OT? Let’s use McCarthy’s idea that *[+RTR, +hi, -back] is responsible for stopping pharyngeal spread.

6. Coming up Thursday (last day)

- Finish talking about features/natural classes if we don’t finish today
- Course wrap-up

References

Alderete, John, Paul Tupper & Stefan A. Frisch. 2013. Phonological constraint induction in a connectionist network: learning OCP-Place constraints from data. *Language Sciences* 37. 52–69. doi:10.1016/j.langsci.2012.10.002.

Boersma, Paul & Joe Pater. 2007. Constructing constraints from language data: the case of Canadian English diphthongs (handout). Paper presented at the 38th meeting of the North East Linguistic Society.

Flack, Kathryn Gilbert. 2007. The Sources of Phonological Markedness. University of Massachusetts, Amherst PhD dissertation.

Flemming, Edward. 2005. Deriving natural classes in phonology. *Lingua* 115(3). 287–309. doi:10.1016/j.lingua.2003.10.005.

Hayes, Bruce. 1999. Phonetically driven phonology: the role of Optimality Theory and inductive grounding. In Michael Darnell, Frederick J Newmeyer, Michael Noonan, Edith Moravcsik & Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*, 243–285. Amsterdam: John Benjamins.

Kirby, James. 2013. The role of probabilistic enhancement in phonologization. In Alan C. L. Yu (ed.), *Origins of sound change*, 228–246. Oxford: Oxford University Press.

Mielke, Jeff. 2004. The emergence of distinctive features. Ohio State University PhD dissertation.

Moreton, Elliott. 2010. Constraint induction and simplicity bias in phonological learning (handout). Workshop on Grammar Induction. Cornell University, ms.

Pater, Joe. 2014. Canadian raising with language-specific weighted constraints. *Language* 90(1). 230–240. doi:10.1353/lan.2014.0009.

Pizzo, Presley. 2013. An online model of constraint induction for learning phonological alternations. Instanbul.