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:

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

\[
(117) \text{Some segment } x \text{ is perceptually difficult in some context } \text{Context}_z \text{ if either:}
\]
\[
a. \quad \text{Accuracy}(x/\text{Context}_z) < \text{threshold}\]
\[
\text{and}
\]
\[
\text{Accuracy}(x/\text{Context}_z) < \text{Accuracy}(y/\text{Context}_z) \rightarrow \text{Constraint } *x/\text{Context}_z
\]
\[
\text{This difference must be significant (} \alpha = 0.01\).
\]
\[
b. \quad \text{Accuracy}(x/\text{Context}_z) < \text{FalseAlarm}(x/\text{Context}_z) \rightarrow \text{Constraint } *x/\text{Context}_z
\]

- Where “accuracy” means hit rate
- If there is no hit rate, because the sound never actually occurs in that context, treat it as 0.
  ▪ 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

![Induced constraints: Summary](image)

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

   ![Diagram showing palatalization rules]

   - 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?

<table>
<thead>
<tr>
<th>Feature System</th>
<th>Characterizable (Natural)</th>
<th>Noncharacterizable (Unnatural)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminaries</td>
<td>3640</td>
<td>2437</td>
</tr>
<tr>
<td>SPE</td>
<td>4313</td>
<td>1764</td>
</tr>
<tr>
<td>Unified Feature Theory</td>
<td>3872</td>
<td>2205</td>
</tr>
<tr>
<td>ANY SYSTEM</td>
<td>4579</td>
<td>1498</td>
</tr>
</tbody>
</table>

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

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

   ![Table showing set operations results]

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

   - 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

<table>
<thead>
<tr>
<th>Best analysis</th>
<th>Preliminaries</th>
<th>SPE</th>
<th>Unified Feature Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural (feature conjunction)</td>
<td>342</td>
<td>467</td>
<td>270</td>
</tr>
<tr>
<td>Disjunction (2 classes)</td>
<td>1718</td>
<td>1994</td>
<td>1745</td>
</tr>
<tr>
<td>Subtraction (2 classes)</td>
<td>9</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Disjunction (3 classes)</td>
<td>948</td>
<td>1160</td>
<td>939</td>
</tr>
<tr>
<td>Disjunction (4 classes)</td>
<td>624</td>
<td>774</td>
<td>630</td>
</tr>
<tr>
<td>Disjunction (5 classes)</td>
<td>349</td>
<td>456</td>
<td>352</td>
</tr>
<tr>
<td>Disjunction (6 classes)</td>
<td>247</td>
<td>292</td>
<td>246</td>
</tr>
<tr>
<td>Disjunction (7 classes)</td>
<td>107</td>
<td>126</td>
<td>121</td>
</tr>
<tr>
<td>Disjunction (8 classes)</td>
<td>29</td>
<td>29</td>
<td>48</td>
</tr>
<tr>
<td>Disjunction (9 classes)</td>
<td>8</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Disjunction (10+) or error</td>
<td>241</td>
<td>290</td>
<td>400</td>
</tr>
<tr>
<td>Unnatural (even w/disjunction)</td>
<td>1455</td>
<td>469</td>
<td>1299</td>
</tr>
</tbody>
</table>

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

• 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):

![Graph showing 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:

![Diagram showing generalization of 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 postnasal 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])

- 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