Overview: What if we aren’t born with a constraint inventory? We’ll take a tour of some proposals.


   - Do we really need to consider such a broad range of constraints?
   - Main idea:
     - When you hear a form (say, [#n…]), compare it to a randomly sampled, hypothetical similar form (e.g., [#ŋ…])
     - If your current grammar assigns better harmony to the sampled neighbor than to the observed form, the grammar needs updating
       - Can you see a problem with this?
     - Refinement: remember the last $n$ forms you heard (here, $n=3000$), and don’t worry about a neighbor that’s in that set
   - How to construct a constraint
     - Make a lattice of all the constraints that could prefer the observed form
       - In the [#n…]/[#ŋ…] example, consider all constraints that penalize [+dorsal]
         - In the feature system used in this paper, [CORONAL] is privative
           - Which means what?
         - Let’s draw some of the lattice.
   - Sample a few of these constraints (here, 3) and add them to the grammar
   - Update weights of this new constraint set
     - Use Perceptron rule to update: similar to Gradual Learning Algorithm, except it cares about the number of violations
   - More precisely: see over
- Tested on English onsets, as in Hayes & Wilson (2006)
- Resulting constraints (over):
  - [___] means “any segment”, i.e. [+segmental]
• Correlation to human judgments of English onsets: -0.829 (vs. Hayes & Wilson’s -0.859)
• Interesting problem: tendency towards over-fitting as learning continues
  - As assessed by worsening correlation to human judgments
See the paper for some ways to improve the model

- Discuss: Both this model and the Hayes/Wilson model are for phonotactic learning. How could we extend either of these to learning alternations?

3. 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]!
  - But they do still have one important piece of information listed above—can you guess what?
The learner’s rule for inducing a constraint:

\[(117) \text{Some segment } x \text{ is perceptually difficult in some context } \text{Context}_Z \text{ if either:}\]

\[\begin{align*}
\text{a. } \text{Accuracy}(x/\text{Context}_Z) &< \text{threshold} \\
\text{and} \\
\text{Accuracy}(x/\text{Context}_Z) &< \text{Accuracy}(y/\text{Context}_Z) \Rightarrow \text{Constraint } *x/\text{Context}_Z \\
\text{This difference must be significant (} \alpha = 0.01). \\
\text{b. } \text{Accuracy}(x/\text{Context}_Z) &< \text{FalseAlarm}(x/\text{Context}_Z) \Rightarrow \text{Constraint } *x/\text{Context}_Z
\end{align*}\]

- 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
4. **A selection of other approaches that we won’t have time to cover in depth**

- **Hayes (1999):** as we saw last time, generate lots of features according to a set of templates, and then select the ones that match the phonetic map well, 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 \[\text{Ɂʌɪs}\] ‘ice’, construct these, among many others:

  (17) **Observed structure**  
  a. D  
  \[-\text{low}\]  
  RAISED Diphthong: A diphthong must be \[-\text{low}\]  
  (Assign a reward of 1 to each diphthong that is \[-\text{low}\])

  b. D C  
  \[-\text{low}\][\text{-vce}\]  
  RAISED/VOICELESS: A diphthong preceding a voiceless consonant must be raised  
  (Assign a reward of 1 to each diphthong that is \[-\text{low}\] that precedes a voiceless consonant)

- **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 way in 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):** Connectionst model of OCP-Place in Arabic roots

5. **Coming up next Tuesday (there is no class this Thursday)**

- Inducing feature and natural classes
References