Class 18: Exemplars II

To do for next time
- Project abstracts due Friday, June 4 (draft is OK).
- Read Blevins & Garrett (Sameer will present).

Pierrehumbert: exemplar dynamics[1]

1. Background

- In exemplar theory, categories are represented mentally as clouds of remembered tokens (projected onto a similarity map) that are typically densest in the middle.
- Highly similar tokens are grouped into a single exemplar (i.e., the similarity map is granular), whose strength is augmented when tokens are added to that "bin" (and, countervailingly, decays over time, so that more recent tokens have more say, especially for low-frequency words, as in the Goldinger experiment).
- An incoming stimulus with is classified (given a category label) according to the number of exemplars from each category that are similar to it, with a weighting in favor of stronger exemplars.

These figures are pasted from the Pierrehumbert reading!!!

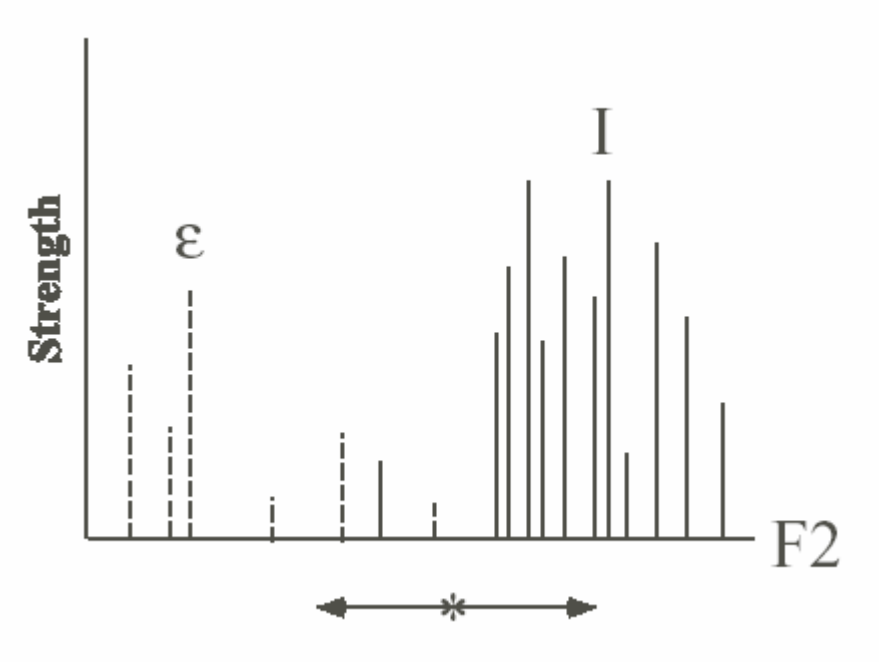


Figure 1, p. 5: the * represents an incoming stimulus's F2 value, the arrows the range of exemplars contributing to the classification decision.

2. Production

To produce an instance of a category, choose an exemplar at random, but with a preference for stronger exemplars. Add some noise, so that the actual production may differ slightly from the exemplar chosen.

[1] Pierrehumbert, Janet (2000). Exemplar dynamics: word frequency, lenition and contrast. In J. Bybee and P. Hopper (eds.), *Frequency effects and emergent grammar.* Amsterdam: John Benjamins.

Pierrehumbert shows that when exemplars are chosen in this way and the resulting tokens added to memory (i.e., to the speaker's memory—if you don't believe in that, I think the model is straightforwardly extended to speaker-listener interactions), the exemplar cloud gradually becomes more diffuse, but its center does not shift.

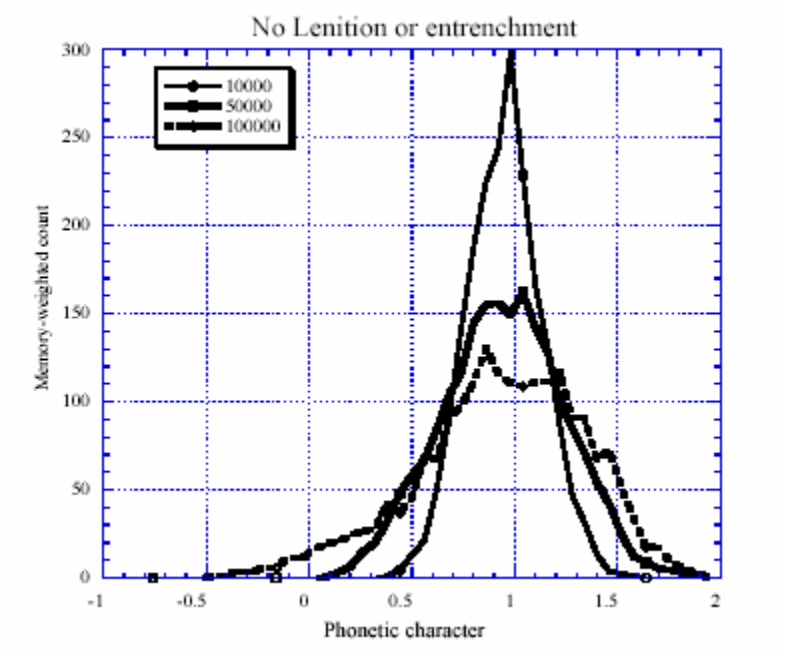


Figure 2, p. 9

3. Lenition

If there is an externally imposed tendency for productions to be slightly hypoarticulated with respect to the exemplar chosen for production, the center of the exemplar cloud gradually shifts towards hypoarticulation (and also becomes more diffuse).

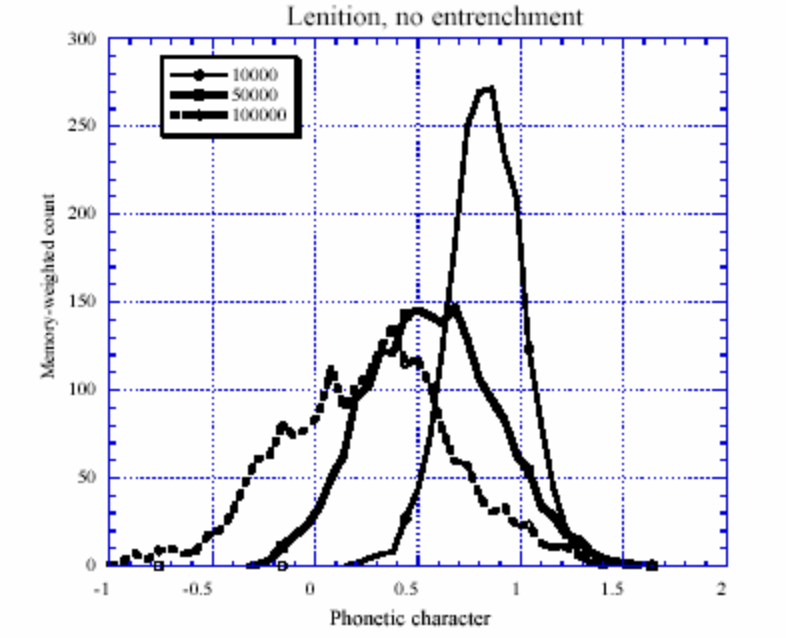


Figure 3, p. 10

For example, if an exemplar is chosen whose remembered articulatory strength along some dimension is 0.9, it may be produced with 0.89 instead. The 0.89 token is then added as an exemplar, and if it is chosen in a later production, it may be pronounced with 0.88, and so on.

In order to avoid diffusion of the exemplar cloud, Pierrehumbert suggests an additional factor in production: instead of picking a (preferentially strong) exemplar, pick a location in the cloud and let all neighboring exemplars contribute in proportion to their strength—this will tend to pull all productions towards the center of the cloud.

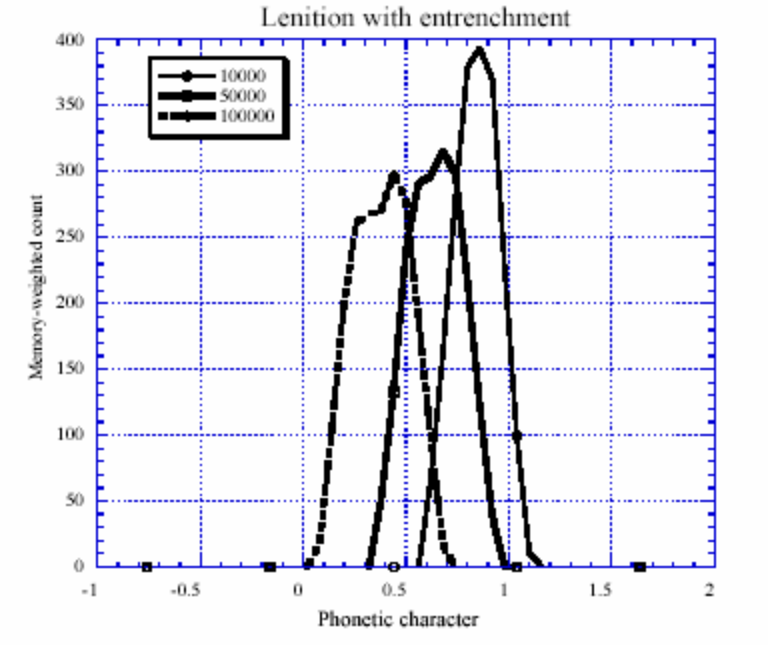


Figure 4, p. 12

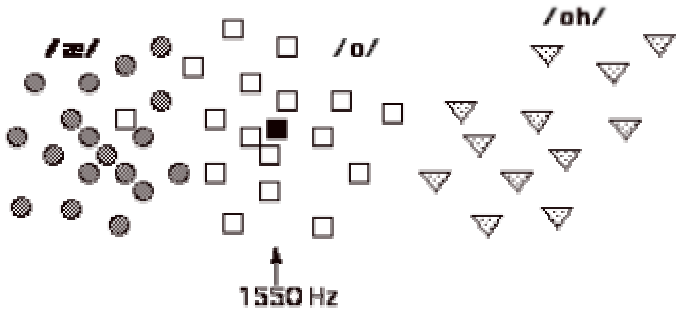
Pierrehumbert leaves open whether...

- all this should happen in production or perception or both.
- the neighborhood to be used should have a fixed "geographic" size or a fixed "population" size.
- feedback (success of the communicative interaction) should play a role in the speaker's updating. (Tabor 1994:[2] you don't need explicit feedback to the speaker if you just exclude unsuccessful communications as learning data for the listener.)

4. Aside: an intriguing proposal from Labov[3]

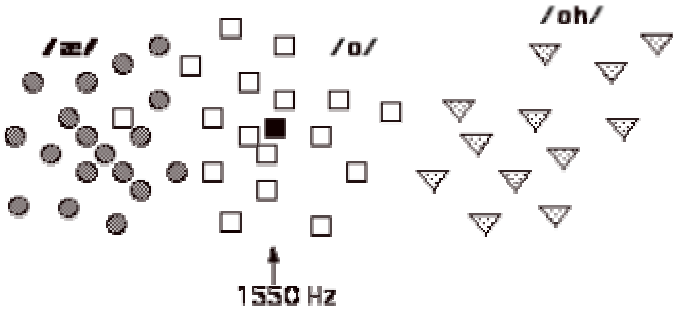
A stable vowel system is, well, stable, because outlying tokens should have a tendency to get misclassified, and thus not contribute to the typical realization of that vowel:

[2] Tabor, Whitney (1994). *Syntactic Innovation: A Connectionist Model*. Stanford University dissertation.
[3] Labov, William (2002). Driving forces in linguistic change. To appear in proceedings of the 2002 International Conference on Korean Linguistics.



"A single aberrant token of /o/ is well within the /æ/ distribution. It may well be identified as /o/, and contribute to the computed mean value of /o/. But the listener may also fail to comprehend the word and the sentence it contains. In that case, it will not be contribute to the mean value of /o/. The end result will be a computed mean for the second formant of F2 as 1550 Hz (the black square)."

But if one vowel is shifting, a neighboring vowel can be pulled along…



"The fronted token of /o/ will no longer lie within the range of the /æ/ distribution, and there is a much greater likelihood that it will be identified correctly as /o/. It will then contribute to the computation of the mean value of /o/, and accordingly, that mean will be shifted towards the front—here shown as 1571 Hz."

Another candidate for simulation to see if this can really work…

5. Back to Pierrehumbert: Frequency

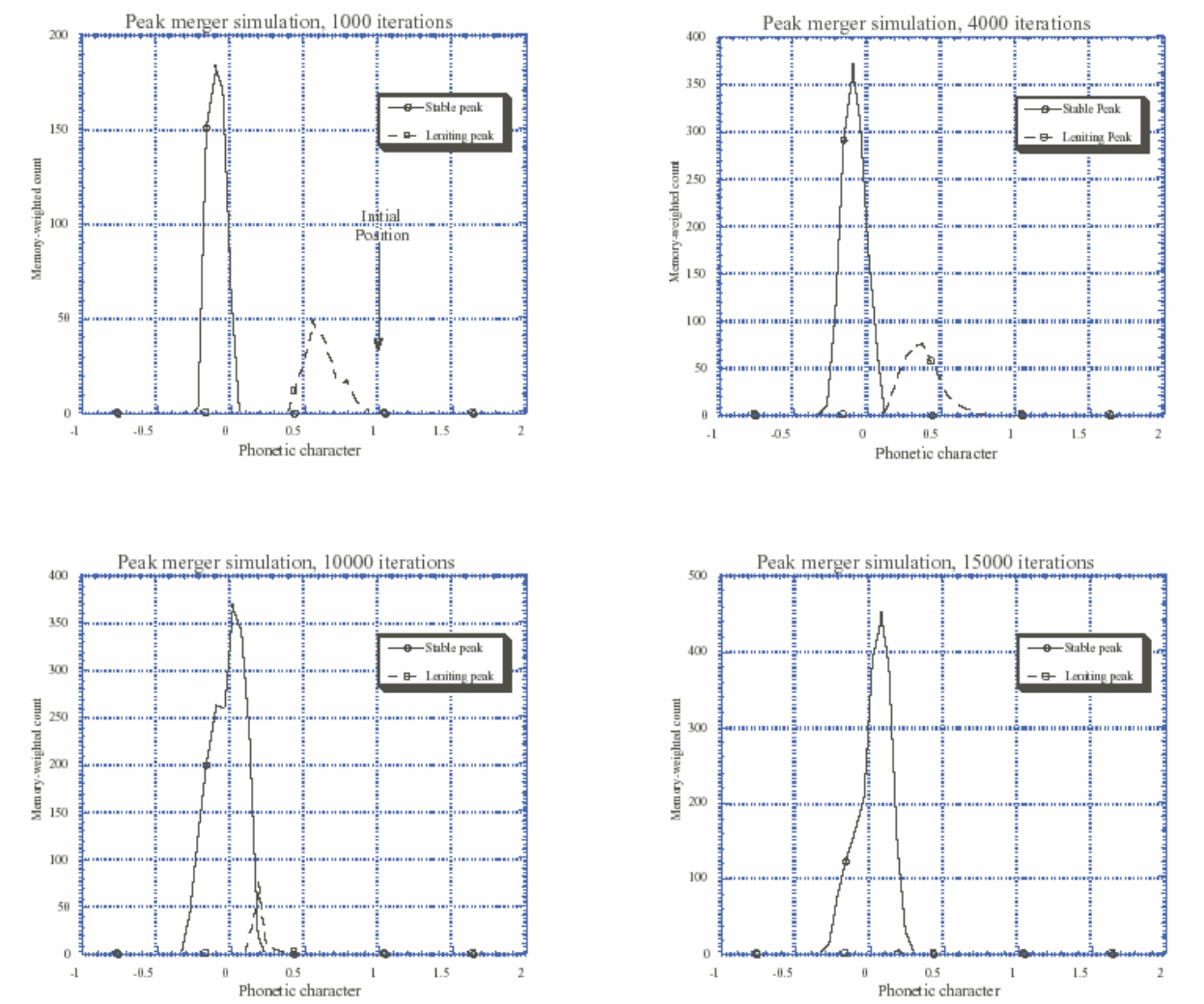
The shift towards lenition increases as the number of productions of the category increases.

This means that if individual words have their own exemplar clouds, then words that are used more often shift more rapidly, as predicted by Bybee. (Figure 3 or 4, interpreted synchronically) Or, in the Pierrehumbert 2002 model, if roughly phoneme-sized tokens are indexed to the words they come from, then the subset of the cloud of tokens that is indexed to a frequent word will shift more quickly.

Pierrehumbert further shows how an infrequent category that is subject to lenition (or any other persistent, unidirectional bias) gets absorbed into a frequent category that is not subject to lenition:

As the less-frequent category (A)'s center moves towards the more-frequent category (B)'s, some tokens of A are miscategorized as B (this happens more often than the reverse because the stronger exemplars of B have more attracting power).

Figure 5, p. 15



This means that neutralizations should be in the direction of the more frequent category. Is this generally true? It should often be true (by accident) if it's the case that less marked categories are more frequent. Can we find cases that tease frequency and markedness apart?

6. Neighborhoods?

- Wright's finding that CVC words with dense neighborhoods have more extreme V formants. The denser the neighborhood, the more difficult recognition, so the more failures of recognition there will be for hypoarticulated tokens → the more the representation is skewed towards hyperarticulated tokens.