24.914

Language Variation and Change

An exemplar-based model of sound change

Readings and assignments

- Meet with me about a final paper topic this week
- Short paper 3: Computational models of sound change

Pierrehumbert (2001): An exemplar-based model of sound change

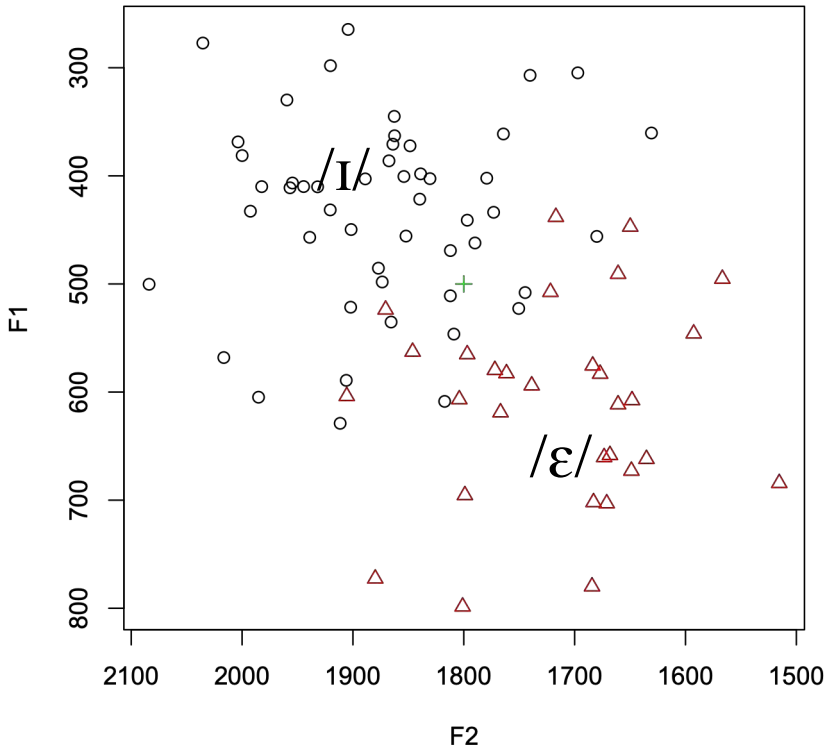
- Pierrehumbert (2001) proposes a partial model of sound change based on exemplar-based models of speech production and perception.
  - Exemplar models: phonetic representations of words/sounds consist of multiple records of utterances of the word/sound
- The main goal of the paper is to account for a putative generalization that sound change proceeds faster in higher frequency words.
  - The core proposal is that change applies to individual words each time they are used, so words that are used more often change faster.

Word frequency and lenition

- Optional schwa deletion is less likely to apply to low frequency words (Hooper 1976)
  - *mammary vs. memory*
- t/d deletion 'is more prevalent in high-frequency words than in low-frequency words' (Bybee 2000)
  - *told* vs. *meant*
- not clear that these are sound changes in progress
- t > ɾ in NZE is a change in progress, and according to Hay & Foulkes (2016) is progressing faster in higher frequency words.

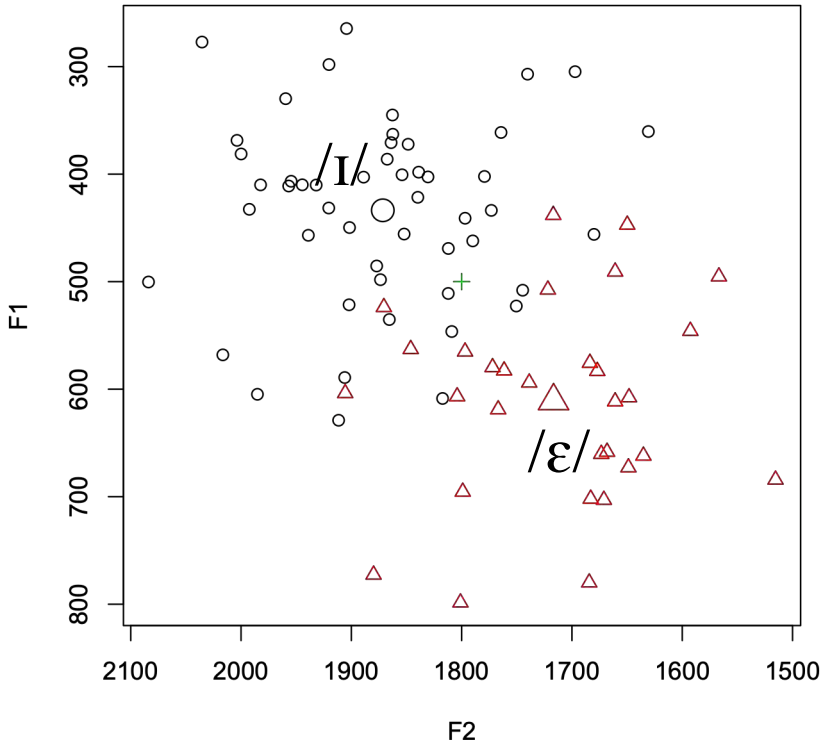
Models of categorization

- We have seen examples of /ɪ/ and /ɛ/ vowels, and we need to categorize a new vowel as an instance of one vowel or the other.



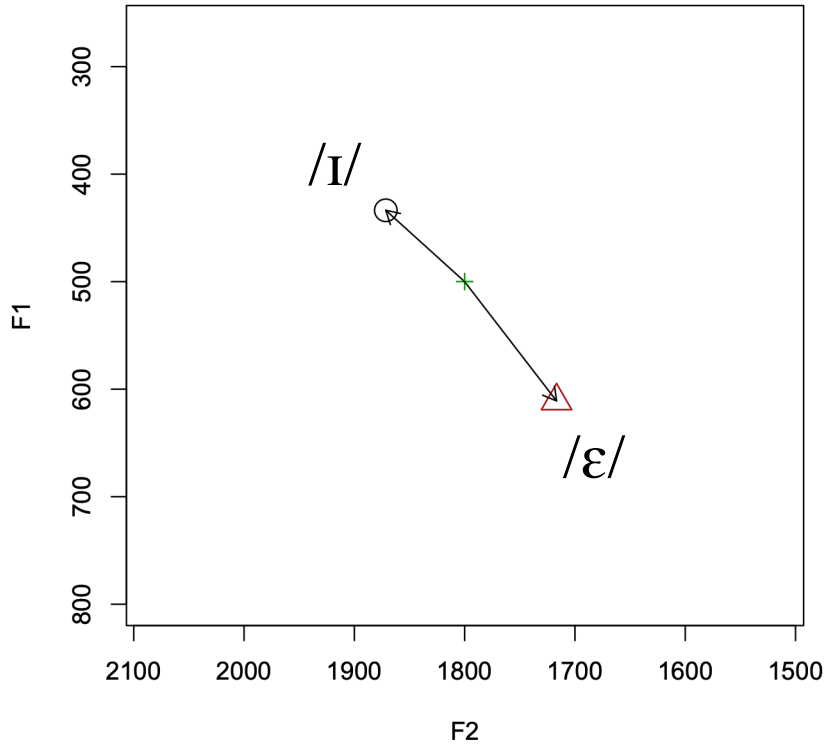
Models of categorization

- We could use the examples of of /ɪ/ and /ɛ/ to learn prototypes of these vowels (e.g. mean formant values)



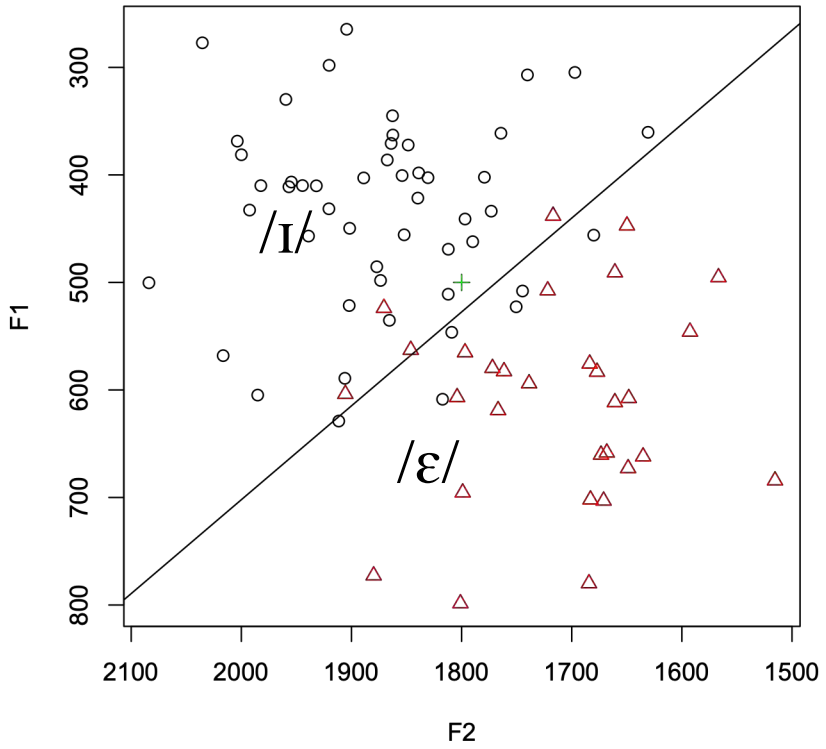
Models of categorization

- We could use the examples of of /ɪ/ and /ɛ/ to learn prototypes of these vowels (e.g. mean formant values)
- Then categorize based on distance to these prototypes



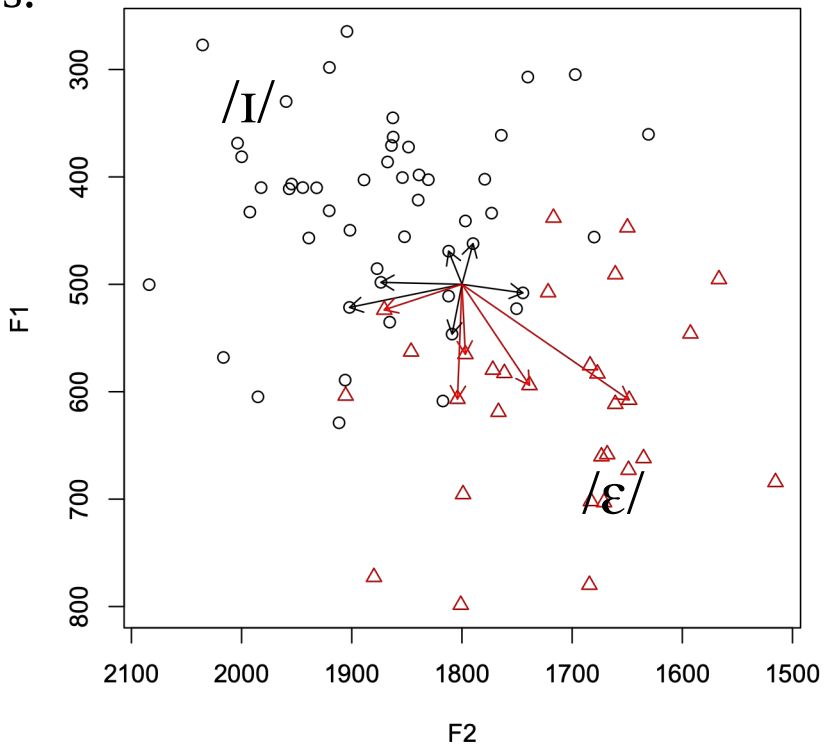
Models of categorization

- Or construct a category boundary based on the learning data, and categorize stimuli based on that.



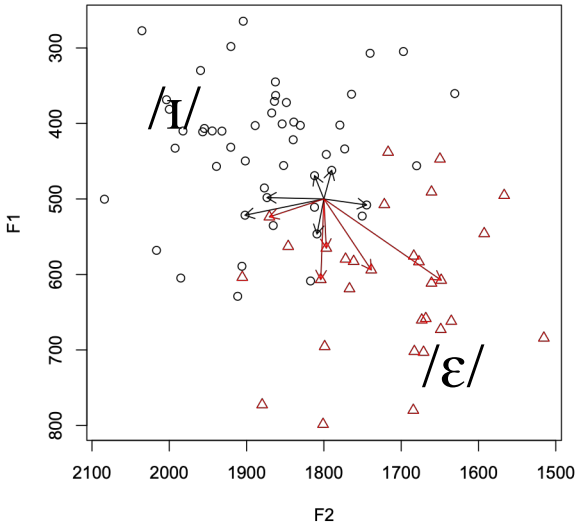
Exemplar model of categorization

- Store all of the learning data (‘exemplars’), then categorize stimuli based on summed similarities to exemplars.
  - Similarity decays exponentially with distance
  - Assign the stimulus to the category with the greatest summed similarities.



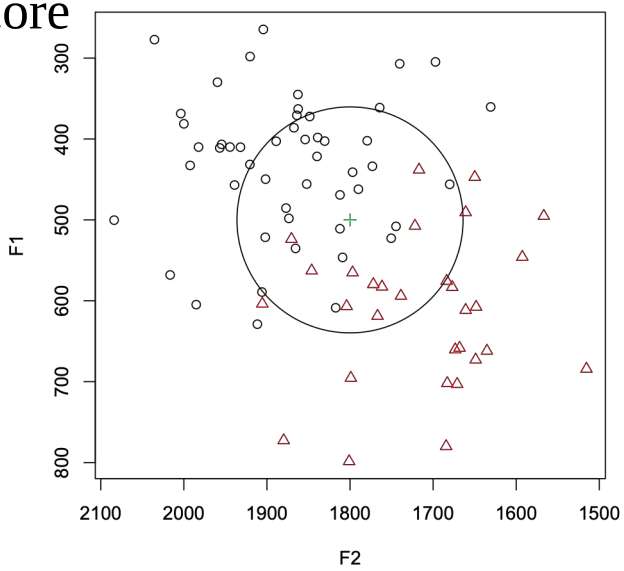
Exemplar model of categorization

- Store all of the learning data ('exemplars'), then categorize stimuli based on summed similarities to exemplars.
- Similarity between *i* and *j*, $\eta_{ij} = e^{-d_{ij}}$
  - where d_{ij} is the euclidian distance between *i* and *j* (Nosofsky 1986)
- Assign the stimulus to the category with the greatest summed similarities.
  - or the probability of assigning stimulus to a category is proportional to the summed similarities to the exemplars of that category



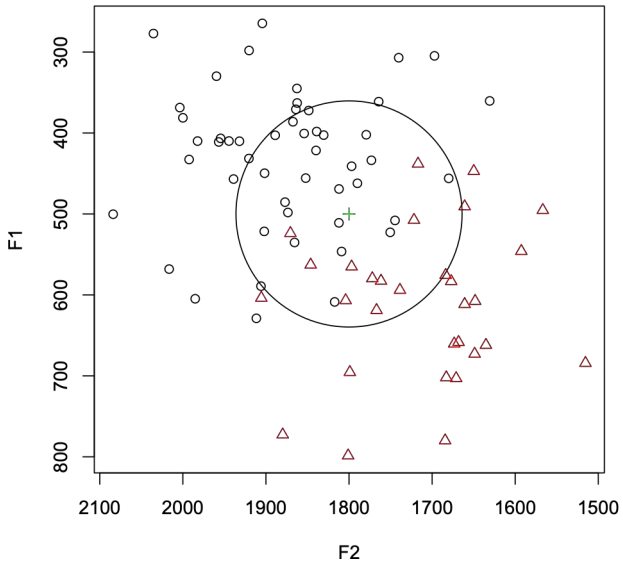
Exemplar model of categorization

- Store all of the learning data (‘exemplars’), then categorize stimuli based on summed similarities to exemplars.
- Or draw a circle around the stimulus and count the number of exemplars of each category in the circle (Pierrehumbert 2001)
  - Assign the stimulus the category with the highest score.
- Exemplar models give an important role to exemplar frequency
  - A category with more exemplars will tend to have a higher similarity score



Exemplar model of categorization

- Exemplar models give an important role to exemplar frequency
  - A category with more exemplars will tend to have a higher similarity score since each exemplar contributes to the score.
- Frequency often does affect categorization, e.g. listeners tend to be more likely to identify a stimulus as a more frequent word.
  - However this effect is context-dependent.
  - Nosofsky posits separate response bias parameters rather than relying on exemplar frequency to derive bias effects.



Exemplar decay

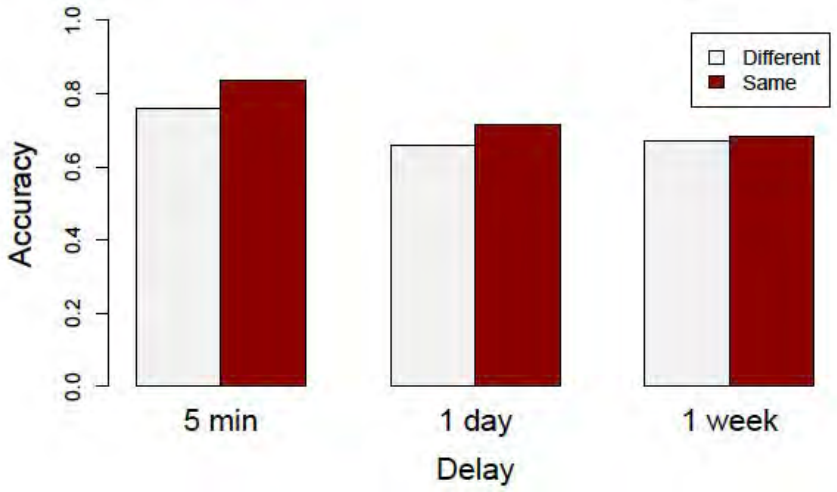
- Exemplars are hypothesized to decay over time.
- Each exemplar has an activation – more active exemplars contribute more to the calculation of similarity to a category.
  - Similarity to an exemplar is weighted by its activation.
  - Activation decays exponentially $e^{-\frac{t}{\tau}}$
- So more recent exemplars play a greater role in defining categories
- Cf. Paul on motory sensations:
  - ‘this sensation is the product of all the earlier impressions received in the course of carrying out movement in question…the motory sensation must be somewhat modified with each new impression’
  - ‘the later impressions always have stronger after-influences than the earlier’

Episodic memory for speech

- Exemplar models of speech perception rely on detailed memory for individual utterances (‘episodic memory’)
- There is direct evidence for this hypothesis,
  - e.g. Goldinger, S.D. (1996) Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:1166–1183.
- Listeners identified spoken words in a ‘study’ session
  - 150 words, had to type words as they heard them
  - words spoken by 2, 6 or 10 different speakers
- Then in test sessions, subjects heard words and had to say whether they had heard them during the study session or not.
  - 300 words
  - ‘old’ words might be spoken in the same voice as in the study session, or in a different voice.
  - test sessions administered after 5 minutes, 1 day, and 1 week.

Goldinger (1996)

- Subjects are more accurate in recognizing previously heard words if they are presented in the same voice.
  - Effect persisted for a day, but could not be detected after a week.
- E.g. accuracy in the two-voice condition
  - No significant effect of number of voices



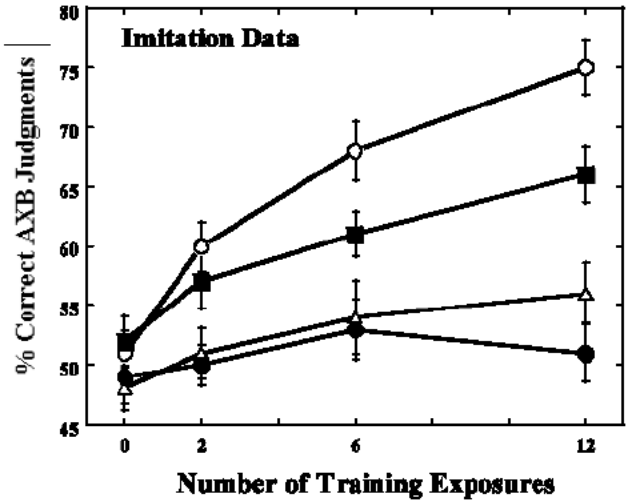
- Also more accurate in identifying words in noise if spoken in the same voice as in the study session
  - This effect was still present after a week.
- Listeners retain ‘detailed episodic traces of spoken words’

Effects of episodic memories on production

- Goldinger (2000) The role of perceptual episodes in lexical processing. In *Proceedings of the Workshop on Spoken Word Access Processes.* MPI for Psycholinguistics, Nijmegen:155–158.
- Day 1: Subjects are recorded reading a list of 160 words
- Day 2: Subjects hear the same words spoken by two male and two female speakers
  - Words were presented 0, 2, 6 or 12 times.
- Day 7: Subjects are recorded reading the same list of 160 words.
- For each word, the renditions from days 1 and 7 were presented together with the recording heard on day 2 to a new set of subjects
  - They had to judge which rendition was more similar to the recording.
- The recordings produced 6 days after the listening session were judged to be more similar to the recording heard in that day.
- Speakers pronunciation was influenced by the pronunciations they had heard 6 days earlier.

Effects of episodic memories on production

- Effect was bigger for lower frequency words

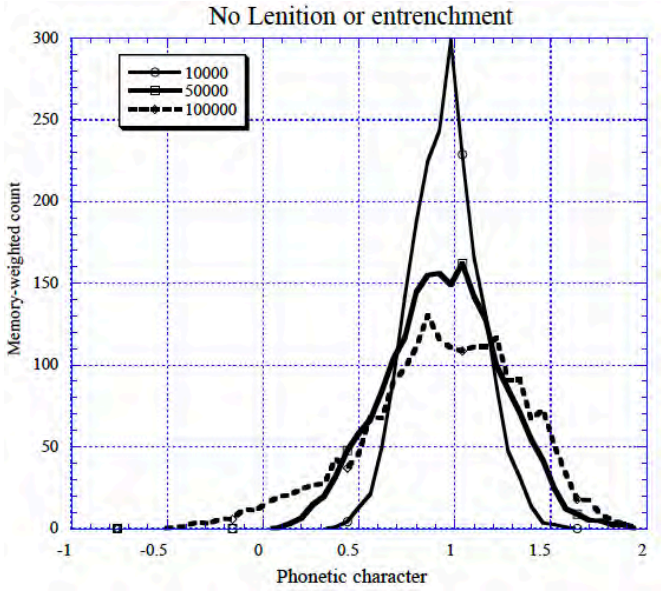


High frequency

Low frequency

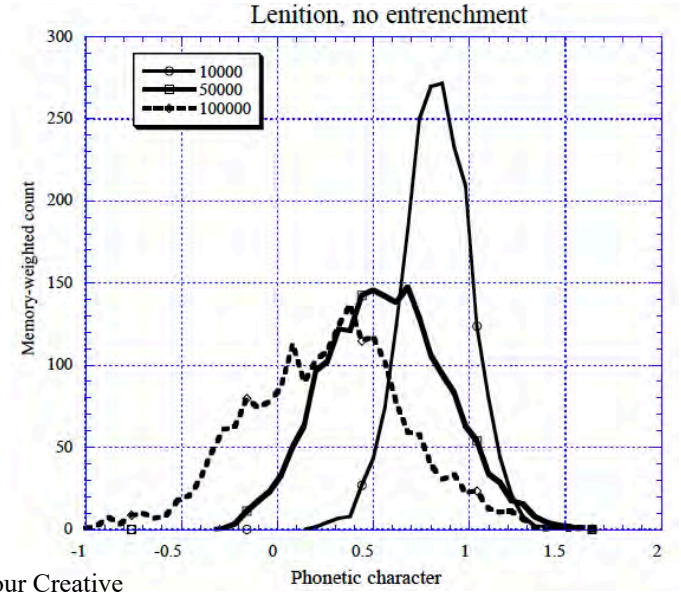
An exemplar model of speech production

- Pierrehumbert proposes a model that allows for such effects
- Model 1: production of a linguistic category (e.g. word) involves selecting an exemplar of that category at random and using it as a model for production.
  - Probability of selection is weighted by exemplar activation
  - Reproduction is imperfect: Noise is added to the selected exemplar.
- Language change via the perception-production loop
  - Start with a single vowel exemplar
  - Produce (with noise, uniform +/-0.1)
  - Store the result
  - Repeat
- Category mean is unchanged
  - unbiased noise
- Variance increases
- Not a good model of acquisition or change



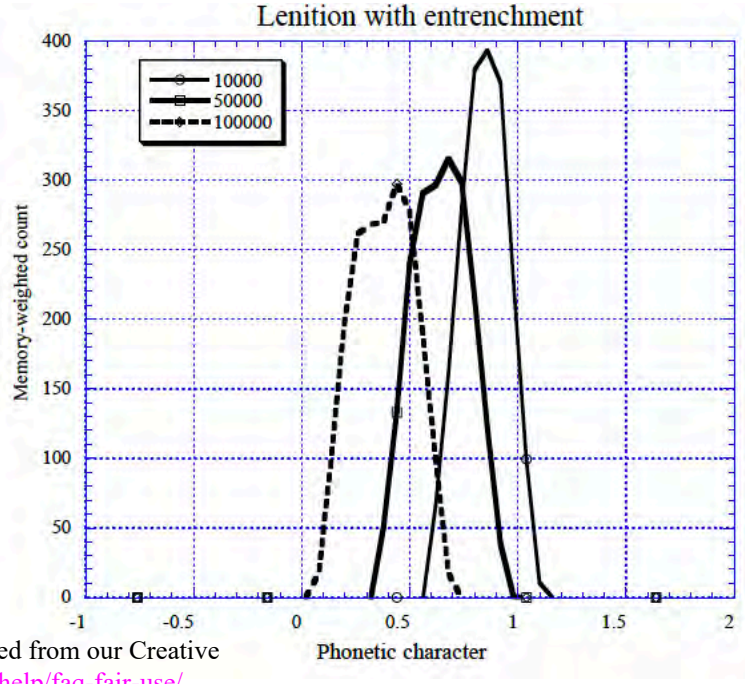
Biased production

- Attributes leniting changes to a bias active in speech production (cf. Paul) – effort reduction.
- When producing a category, a bias (-0.01) is added to the selected exemplar in addition to noise
  - Note: No attempt to address the actuation problem (when/why such a bias takes effect).
- Results in change in category mean and increase in variance
- If exemplars are words rather than phonemes then this model predicts that leniting changes apply faster in more frequent words.
  - bias applies each time a word is produced
  - more frequent words are affected more often.



Entrenchment

- A modification is proposed to oppose the tendency for category variance to increase over time:
- Production involves selecting a set of exemplars and averaging them
  - 500 exemplars
  - averaging weighted by activation
- Averaging pulls new exemplars towards the mean of the distribution
- Lenition bias plus entrenchment:



What are the exemplars?

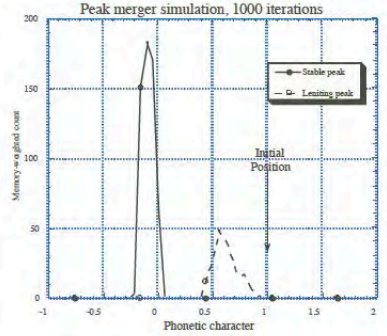
- Words? Phonemes? Other?
- What exemplars are required to model the frequency effect?
- How about regular change?
  - Can a change go to completion in this model?
- Conditioned change (e.g. oʊ > oʊ/_ɬ, oʊ > əʊ elsewhere)?

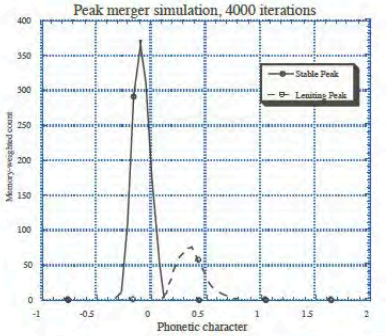
- Are exemplars essential to the model? What role do they play?

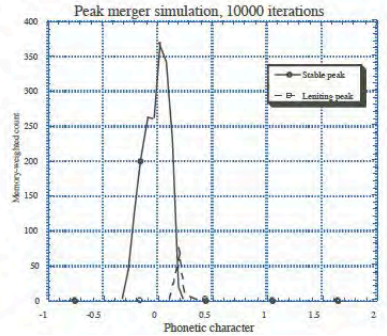
Neutralization

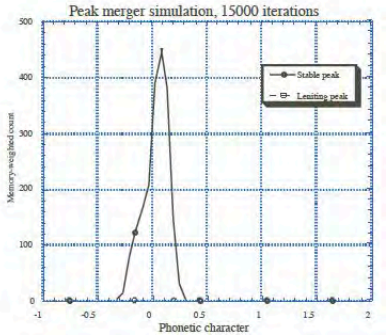
- Neutralization of two categories is presented as a way in which regular change can arise – i.e. all relevant words end up with the same realization.
- Two categories:
  - One subject to bias, the other not.
  - The fixed category has higher frequency (3×)

- Can this model derive regular sound change without neutralization?









24.914 Language Variation and Change
Spring 2019